

71

studies in physical and theoretical chemistry

1963
15443
COPY

1

AD-A225 805



71

DTIC
ELECTE
AUG 23 1990
S B D

DISTRIBUTION STATEMENT A
Approved for public release;
distribution is unlimited

90 08 17 063

studies in physical and theoretical chemistry 71

**MODELLING OF MOLECULAR
STRUCTURE AND PROPERTIES**

1
studies in physical and theoretical chemistry 71

MODELLING OF MOLECULAR STRUCTURES AND PROPERTIES

Proceedings of the 44th International Meeting of Physical Chemistry on
Modeling of Molecular Structures and Properties in Physical Chemistry
and Biophysics, Organised by The Division de Chimie physique of the
Société Française de Chimie, Nancy, France, 11–15 September 1989

Edited by

J.-L. RIVAIL

*Laboratoire de Chimie théorique (BP 239),
54506 Vandoeuvre-Les-Nancy,
France*

Co-editors

J.Y. LALLEMAND
R. LAVERY
J.P. MORNON
G. PEPE
B. ROBINET
E. SOULIE
C. TROYANOWSKY
G. VERGOTEN

DTIC
ELECTE
AUG 23 1990
S B D



ELSEVIER

Amsterdam — Oxford — New York — Tokyo 1990

DISTRIBUTION STATEMENT A

Approved for public release;
Distribution Unlimited

ELSEVIER SCIENCE PUBLISHERS B.V.
Sara Burgerhartstraat 25
P.O. Box 211, 1000 AE Amsterdam, The Netherlands

Distributors for the United States and Canada:

ELSEVIER SCIENCE PUBLISHING COMPANY INC.
655, Avenue of the Americas
New York, NY 10010, U.S.A.

Library of Congress Cataloging-in-Publication Data

Société française de chimie. Division de chimie physique.
International Meeting (44th : 1989 : Nancy, France)

Modelling of molecular structures and properties : proceedings of
the 44th International Meeting of Physical Chemistry of modelling of
molecular structures and properties in physical chemistry and
biophysics / organized by the Division de chimie physique of the
Société française de chimie, Nancy, France, 11-15 September 1989 :
edited by J.L. Rivail.

p. cm. -- (Studies in physical and theoretical chemistry :
71)

Includes bibliographical references.

ISBN 0-444-88714-8

1. Molecules--Models--Congresses. 2. Molecular structure--
Congresses. 3. Chemical bonds--Models--Congresses. I. Rivail, J.
L. (Jean Louis) II. Société française de chimie. Division de
chimie physique. III. Title. IV. Series.

QD480.S63 1989

541.2'2--dc20

90-3270
CIP

ISBN 0-444-88714-8

© Elsevier Science Publishers B.V., 1990

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior written permission of the publisher, Elsevier Science Publishers B.V./Physical Sciences & Engineering Division, P.O. Box 330, 1000 AH Amsterdam, The Netherlands.

Special regulations for readers in the USA - This publication has been registered with the Copyright Clearance Center Inc. (CCC), Salem, Massachusetts. Information can be obtained from the CCC about conditions under which photocopies of parts of this publication may be made in the USA. All other copyright questions, including photocopying outside of the USA, should be referred to the copyright owner, Elsevier Science Publishers B.V., unless otherwise specified.

No responsibility is assumed by the Publisher for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions or ideas contained in the material herein.

This book is printed on acid-free paper.

Printed in The Netherlands

AVAILABLE FOR \$269.25 from Elsevier
Science Publishing Co., Inc.
52 Vanderbilt Ave., New York, NY 10017
TELECON 8/22/90 VG

Price \$ 269.25

per telecon

AVAILABILITY CODES

Avail and/or Special	
Dist	
A-1	21



studies in physical and theoretical chemistry

Other titles in this series

- 11 **Electrodes of Conductive Metallic Oxides. Part A** edited by S. Trasatti
Electrodes of Conductive Metallic Oxides. Part B edited by S. Trasatti
- 12 **Ionic Hydration in Chemistry and Biophysics** by B.E. Conway
- 13 **Diffraction Studies on Non-Crystalline Substances** edited by I. Hargittai and W.J. Orville-Thomas
- 14 **Radiation Chemistry of Hydrocarbons** by G. Földiák
- 15 **Progress in Electrochemistry** edited by D.A.J. Rand, G.P. Power and I.M. Ritchie
- 16 **Data Processing in Chemistry** edited by Z. Hippe
- 17 **Molecular Vibrational-Rotational Spectra** by D. Papoušek and M.R. Aliev
- 18 **Steric Effects in Biomolecules** edited by G. Náray-Szabó
- 19 **Field Theoretical Methods in Chemical Physics** by R. Paul
- 20 **Vibrational Intensities in Infrared and Raman Spectroscopy** edited by W.B. Person and G. Zerbi
- 21 **Current Aspects of Quantum Chemistry 1981** edited by R. Carbó
- 22 **Spin Polarization and Magnetic Effects in Radical Reactions** edited by Yu.N. Molin
- 23 **Symmetries and Properties of Non-Rigid Molecules: A Comprehensive Survey** edited by J. Maruani and J. Serre
- 24 **Physical Chemistry of Transmembrane Ion Motions** edited by G. Spach
- 25 **Advances in Mössbauer Spectroscopy: Applications to Physics, Chemistry and Biology** edited by B.V. Thosar and P.K. Iyengar
- 26 **Aggregation Processes in Solution** edited by E. Wyn-Jones and J. Gormally
- 27 **Ions and Molecules in Solution** edited by N. Tanaka, H. Ohtaki and R. Tamamushi
- 28 **Chemical Applications of Topology and Graph Theory** edited by R.B. King
- 29 **Electronic and Molecular Structure of Electrode-Electrolyte Interfaces** edited by W.N. Hansen, D.M. Kolb and D.W. Lynch
- 30 **Fourier Transform NMR Spectroscopy (second edition)** by D. Shaw
- 31 **Hot Atom Chemistry: Recent Trends and Applications in the Physical and Life Sciences and Technology** edited by T. Matsuura
- 32 **Physical Chemistry of the Solid State: Applications to Metals and their Compounds** edited by P. Lacombe
- 33 **Inorganic Electronic Spectroscopy (second edition)** by A.B.P. Lever
- 34 **Electrochemistry: The Interfacing Science** edited by D.A.J. Rand and A.M. Bond
- 35 **Photophysics and Photochemistry above 6 eV** edited by F. Lahmani
- 36 **Biomolecules: Electronic Aspects** edited by C. Nagata, M. Hatano, J. Tanaka and H. Suzuki
- 37 **Topics in Molecular Interactions** edited by W.J. Orville-Thomas, H. Ratajczak and C.N.R. Rao
- 38 **The Chemical Physics of Solvation. Part A. Theory of Solvation** edited by R.R. Dogonadze, E. Kálmán, A.A. Kornyshev and J. Ulstrup
The Chemical Physics of Solvation. Part B. Spectroscopy of Solvation edited by R.R. Dogonadze, E. Kálmán, A.A. Kornyshev and J. Ulstrup
The Chemical Physics of Solvation. Part C. Solvation Phenomena in Specific Physical, Chemical, and Biological Systems edited by R.R. Dogonadze, E. Kálmán, A.A. Kornyshev and J. Ulstrup
- 39 **Industrial Application of Radioisotopes** edited by G. Földiák

- 40 **Stable Gas-in-Liquid Emulsions: Production in Natural Waters and Artificial Media** by J.S. D'Arrigo
- 41 **Theoretical Chemistry of Biological Systems** edited by G. Náray-Szabó
- 42 **Theory of Molecular Interactions** by I.G. Kaplan
- 43 **Fluctuations, Diffusion and Spin Relaxation** by R. Lenk
- 44 **The Unitary Group in Quantum Chemistry** by F.A. Matsen and R. Pauncz
- 45 **Laser Scattering Spectroscopy of Biological Objects** edited by J. Štěpánek, P. Anzenbacher and B. Sedláček
- 46 **Dynamics of Molecular Crystals** edited by J. Lascombe
- 47 **Kinetics of Electrochemical Metal Dissolution** by L. Kiss
- 48 **Fundamentals of Diffusion Bonding** edited by Y. Ishida
- 49 **Metallic Superlattices: Artificially Structured Materials** edited by T. Shinjo and T. Takada
- 50 **Photoelectrochemical Solar Cells** edited by K.S.V. Santhanam and M. Sharon
- 51 **Graph Theory and Topology in Chemistry** edited by R.B. King and D.H. Rouvray
- 52 **Intermolecular Complexes** by P. Hobza and R. Zahradník
- 53 **Potential Energy Hypersurfaces** by P.G. Mezey
- 54 **Math/Chem/Comp 1987** edited by R.C. Lacher
- 55 **Semiconductor Electrodes** edited by H.O. Finklea
- 56 **Computational Chemistry** by M.D. Johnston, Jr.
- 57 **Positron and Positronium Chemistry** edited by D.M. Schrader and Y.C. Jean
- 58 **Ab Initio Calculation of the Structures and Properties of Molecules** by C.E. Dykstra
- 59 **Physical Adsorption on Heterogeneous Solids** by M. Jaroniec and R. Madey
- 60 **Ignition of Solids** by V.N. Vilyunov and V.E. Zarko
- 61 **Nuclear Measurements in Industry** by S. Rózsa
- 62 **Quantum Chemistry: Basic Aspects, Actual Trends** edited by R. Carbó
- 63 **MATH/CHEM/COMP 1988** edited by A. Graovac
- 64 **Valence Bond Theory and Chemical Structure** edited by D.J. Klein and N. Trinajstić
- 65 **Structure and Reactivity in Reverse Micelles** edited by M.P. Pileni
- 66 **Applications of Time-Resolved Optical Spectroscopy** by V. Brückner, K.-H. Feller and U.-W. Grummt
- 67 **Magnetic Resonance and Related Phenomena** edited by J. Stankowski, N. Pislewski, S.K. Hoffmann and S. Idziak
- 68 **Atomic and Molecular Clusters** edited by E.R. Bernstein
- 69 **Structure and Properties of Molecular Crystals** edited by M. Pierrot
- 70 **Self-Consistent Field: Theory and Applications** edited by R. Carbó and M. Klobukowski

CONTENTS

Foreword	xii
Avant propos	xiv
Acknowledgements	xvi
List of participants	xvii

Molecular modelling, with or without quantum chemistry? (B. Pullman) ...	1
--	---

I. THEORETICAL ADVANCES AND BASIC METHODOLOGY

General theory of intermolecular forces (A.D. Buckingham)	17
Theoretical models for intermolecular potentials (A.J. Stone)	27
A vibrational molecular force field of compounds with biological interest (M. Dauchez, P. Derreumaux, P. Lagant and G. Vergoten)	45
An introduction to numerical optimization (C. Lemarechal)	63
Describing protein conformation: a new mathematical approach K. Zakrzewska and R. Lavery)	81
GENMOL. A fast program for molecular modeling. Application to the determination of the psychotonic or sedative effects of tricyclic antidepressant drugs (G. Pepe and D. Siri)	93

II. MOLECULAR MODELLING IN CHEMISTRY

IIa. Catalysis and adsorption

The free energy of intercalation: the structure of graphite intercalation compounds (Z.-M. Chen, O.A. Karim and B.M. Pettitt) ..	103
Chemisorption on transition metal surfaces and heterogeneous catalysis: back-of-the-envelope theoretical modeling (E. Shustorovich)	119
Potential maps of CH ₄ , H ₂ O and CH ₃ OH in silicalite. Influence of the silicalite structure (F. Vigné-Maeder)	135

IIb. Conventional methods

Conformational studies on macrocyclic receptors and on their substrate complexes (G. Wipff)	143
Calculation of the free energy barrier for the electron transfer reaction between Fe ⁺ and Fe ²⁺ in water (A. González-Lafont, J.M. Lluch, A. Oliva and J. Bertran)	165
Design of new efficient radical clocks with the help of molecular mechanics (J.-M. Mattalia, M. Fathallah, A. Samat, B. Blaive and M. Chanon)	173
Phase behaviour of cyclohexane derived from differential thermal analysis and computed by molecular dynamics (C. Hoheisel, M. Hartmann and A. Würflinger)	181
The Na ⁺ ion solvation in water-hexamethylphosphorictriamide mixture: molecular dynamics simulation (I.I. Vaisman, M.G. Kiselev, Y.P. Puhovski and Y.M. Kessler)	187

VIII

Simulation of phase transitions in organic crystals - the gamma-alpha phase transition of p-dichlorobenzene (J. Royer, F. Bayard, C. Decoret and M. Skalli)	195
Monte Carlo simulations of the acetate-guanidinium ion pair in water (S. Boudon and G. Wipff)	203
Simulation of polyionic interactions by progressive approximations (H.H. Ohlenbusch)	211
High temperature annealed molecular dynamics simulations as a tool for conformational sampling: a test case on the 222 bicyclic cryptand (P. Auffinger and G. Wipff)	219
Structural and charge effects on the phthalocyanine dimer (F. Torrens, E. Orti and J. Sanchez-Marin)	221
Structural comparative study of hydroxy and methoxy flavonoids using MNDO, MINDO/3 and AM1 programs (P. Bonifassi, J.C. Wallet and E.M. Gaydou)	229

IIc. Conventional methods applied to biological model systems

Theoretical study of the structure of the glutathione-hydrogen peroxide complex (J. Berges, J. Caillet, J. Langlet, Z. Abedinzadeh and M. Gardes-Albert)	233
Structure and dynamics of water and ionic solutions near biomembrane surfaces from molecular dynamics simulations (M. Schlenkrich, K. Nicklas, J. Böcker, P. Bopp and J. Brickmann)	245
Conformational study of the polar head of the amphotericin B in the isolated state and in the presence of water molecules (J. Berges, J. Caillet, J. Langlet, N. Gresh, M. Herve and C. Gary-Bobo)	253
Transport in biological membranes. Modeling and experiments (E. Garcin, F. Baros, J.C. Andre, D. Daveloose, J. Viret, M.L. Viriot and M. Donner)	265
Relationship between hydrogen bonding properties of some photosystem II herbicides and their activity on triazine-resistant and susceptible chloroplasts (S. Creuzet, B. Gilquin and J.M. Ducruet)	273

IIId. Applications to pharmacology

3D-Pattern recognition of pharmacophors. Application of the method to the comparison morphine-enkephaline (C. Riche and L. Moy)	281
Molecular dynamics simulation of an n-methyl inhibitor bacterial collagenase (B. Gilquin, D. Perahia and V. Dive)	289
Molecular design of new bioactive microproteins from the knotted structure of the trypsin inhibitor EETI II (D. Le Nguyen, A. Heitz, L. Chiche, D. Nalis, B. Castro, A. Favel, H. Mattras, M.A. Coletti-Previero, C. Gaboriau and J.P. Mornon)	297

IIe. Quantum models

Intermolecular forces in some ionic atmospheric clusters involving water. Quantum mechanical and Monte Carlo calculations (E. Kochanski, A. Rahmouni and R. Wiest)	303
Modeling of organometallic reactivity using a combination of extended Hückel and molecular graphics techniques (P.-Y. Morgantini, P. Fluekiger, J. Weber and E.P. Kündig)	313

The water dimer - comparison of results obtained by both ab initio supermolecule and SAPT methods. Derivation of simplified formulas (O. Hess, M. Caffarel, J. Langlet, J. Caillet, C. Huiszoon and P. Claverie)	323
Water molecule in different crystal surroundings: periodic Hartree-Fock and model Hamiltonian calculations (J.G. Ángyán and B. Silvi)	337
Theoretical study of intermolecular proton transfer in glycine (J. Langlet, J. Caillet, E. Evleth and E. Kassab)	345
Double proton transfer studies in carboxylic acid dimers (H. Chojnacki and M.J. Pyka)	351
Modelling of the disulfide bridge in proteins: Ab initio-CI studies of S_2H_2 and $S_2(CH_3)_2$ (M. Loos)	359

IIf. Spectra

Investigation of ionic solvation dynamics by far infrared spectroscopy (B. Guillot, P. Marteau and J. Obriot)	363
DNA structural parameter determination by vibrational data (M. Ghomi) .	373
Force field calculations of siderophores. Stability and conformations of Fe^{III} chelates with novel iron related parameters (A. Bouraoui, M. Fathallah, F. M'Henni, B. Blaive and R. Gallo)	381
Relation entre constantes de force et constantes elastiques dans divers materiaux (E. Husson, Y. Repelin et D. Weigel)	395
A simple model of nucleic bases association: the cyclic dimer of the 2-pyridone vibrational study (J. Favrot, Th. Bouissou, J.F. Brazier, S. Castillo and A. Zwick)	401

IIf. Kinetics

Kinetic modelling of heterogeneous-catalyzed reactions with the AnaCin software. Application to the hydrodenitrogenation of phenanthridine (J. Joffre, P. Geneste, A. Guida, G. Szabo and C. Moreau)	409
Diffusion-controlled reactions. Simulation and new theoretical approach (W. Dong, F. Baros and J.C. André)	417

III. MOLECULAR MODELLING IN BIOPHYSICS

IIIa. Methods

Molecular dynamics: applications to proteins (M. Karplus)	427
On testing theoretical models by comparison of calculated with experimental data (W.F. van Gunsteren)	463
Theoretical study of a conformational change in an enzyme: methodology and first results on citrate synthase (J. Durup)	479
Nonarchimedean modelling for biosystems evolving on many time scales (O. Iordache and P.T. Frangopol)	483

IIIb. Proteins, polypeptides

Molecular dynamics simulations of a solvated protein: analysis of structural and electrostatic properties (S.J. Wodak, D. Van Belle, M. Froeyen and M. Prévost)	491
---	-----

X

Electrostatic free energy as the fundamental structure function correlator in proteins: some perspectives from microscopic simulations of proteins functions (A. Warshel)	515
Modelling the acetylcholine receptor (AChR) channel. Energy profiles and point mutations (A. Pullman, S. Furois-Corbin and A.M. Andrade)	527
Analysing and modelling the deformation of protein secondary structures (C. Etchebest)	541
Monte Carlo free energy calculations in conformational statistics of polypeptide chains (S. Premilat)	549
Evaluation of electrostatic properties at enzyme active sites (G. Dive, J. Lamotte-Brasseur and D. Dehareng)	557
Inhibitor and substrate binding in dihydrofolate reductase - a theoretical approach (T. Šolmajer and M. Hodošček)	567

IIIc. DNA

Modelling base sequence effects and transitions in DNA (R. Lavery, M. Poncin and H. Sklenar)	575
Modelling ionic effects on biomolecular conformation (D.M. Soumpasis) ..	597
Influence of the base sequence and conformation on the structure of the polynucleotide hydration shell (V.I. Poltev, A.V. Teplukhin and G.G. Malenkov)	607
Theoretical prediction of base pair sequence effects in Z-DNA and Z-RNA (B. Hartmann and R. Lavery)	617
Minimization and molecular dynamics of Z-DNA modified by acetylaminofluorene (V. Fritsch and E. Westhof)	627
A theoretical investigation of the base sequence preferences of mono- intercalating polymethylene carboxamide derivatives of 9-aminoacridine (C. Coulombeau and N. Gresh)	635
NMR study and molecular modelling of two double stranded DNA fragments TGACGTCA and ACTGCAGT (J.M. Piriou, O. Mauffret, J. Armier, S. Fermandjian and M. Le Bret)	639
Nucleotide mispairs stabilized by water bridges. Modeling of structure and role in template biosynthesis (V.I. Poltev and S.V. Steinberg) ..	645
Study on the sequence selectivity of the bi-functional intercalator ditercalinium (R.C. Maroun and B.P. Roques)	655

IV. MODELLING-NMR JOINT STUDY OF BIOMOLECULES

NMR studies of proteins and protein-DNA interactions (J.A.C. Rullmann, R. Boelens, R.M.J.N. Lamerichs, G.W. Vuister and R. Kaptein)	661
Restrained molecular mechanics of enkephalin using distances derived from NMR (transferred NOE) (A. Milon)	679
Preliminary structure determination of two toxins using NMR data (A. Mikou, S. Laplante, R. Le Goas, M.-A. Delsuc, I. Charpentier, E. Guittet and J.-Y. Lallemand)	685
Bis-intercalation of ditercalinium and "flexible" analogue in the octanucleotide d(TpTpCpGpCpGgApA) ₂ : a comparative study by NMR and molecular modelling (J. Pothier, M. Delepierre, M.-C. Barsi, C. Garbay, J. Igolen, M. Le Bret and B.P. Roques)	695

Structure determination from NMR using a relaxation matrix approach; application to the solution structure of crambin (J.A.C. Rullmann, R.M.J.N. Lamerichs, C. Gonzalez, T.M.G. Koning, R. Boelens, and R. Kaptein)	703
Combined use of NMR restrained computer simulations for the modeling of the cyclic antifungal lipopeptide: stendomycine (M. Genest, B. Stawarz, J.P. Simorre, D. Genest, A. Caille and M. Ptak)	711
Molecular dynamics studies of pristinamycin II _A : from crystal structure to conformation in apolar solvent using NMR data (E. Surcouf, I. Morize, D. Frechet, M. Vuilhorgne, A. Mikou, E. Guitet and J.Y. Lallemand)	719

V. COMPUTATIONAL SCIENCE

Va. Artificial intelligence

Artificial intelligence in the molecular world (M.C. Haton)	727
An aid to the classification of molecules: presentation of an expert system project (M.P. Gingold)	743
Handling the structural information resulting from molecular modelling for drug activity predictions: the SARAH system (R. Rozot, J.-L. Rivail and H. Mathis)	747
OCTANG system: a new approach of 3D molecular geometry (E. Delance, J.P. Doucet and J.E. Dubois)	755

Vb. Molecular graphics

Molecular graphics for the Macintosh (J.M. Cense)	763
Presentation of GLOS (G.A. Langlet)	767
Presentation of BEMOL (G.A. Langlet)	769

Vc. Computer architecture

Network computing of the Epstein-Nesbet second-order correction to the molecular electronic density. Preliminary results (M. Picard and J.M. Leclercq)	771
Computational chemistry on supercomputers. Introduction to the panel discussion on supercomputing in chemistry (F. Schwaab)	777

FOREWORD

The improved techniques of molecular graphics have triggered very fast advances in molecular modelling, be it concerned with structures or behaviour. Jean-Louis Rivail's proposal to devote our international meeting to molecular modelling was therefore approved practically without discussion. The way our conference went proved it had been a good choice.

May I recall here what is nearly a return to Genesis. I have a vivid memory, as a student, of attending a symposium on "The Chemical Bond" held in 1948 in Paris: Mulliken, Pauling, Raman, Louis de Broglie, Ketelaar, Cabannes, Coulson, and the bright young people of theoretical chemistry compared views that - fortunately - did not always converge. This symposium is now practically part of the legend of theoretical chemistry, at a time when many experimentalists were judging it an intellectual game of dubious interest: "I'll believe in these computations when those people will be able to predict something, not only of demonstrating what is already known." Theoreticians have since proved themselves.

They were fabulously helped by the daily improvement in computer performance. As a young research fellow I remember being shown, in the research lab. of a great hardware producer, a computer occupying nearly 100 square metres of floor, and which could compute "so fast" that its innumerable errors were unimportant: it was sufficient to do the same computation 100 times and plot the corresponding Gaussian curve: its summit gave the right value. One dreams at such memories.

At the same time crystal structures compelled very lengthy work, whereas nowadays one must select, because of its interest, a structure which will be determined very quickly. Conversely the field of "chemical computation" is still shared between some research on molecular reactivity that involve laborious "number crunching", while the field of molecular modelling offers many subtle examples of software which can visualize simply enough the structures and behaviours of complex molecules.

Our Scientific Committee - J.Y.Lallemand, R.Lavery, J.P.Mornon, G.Pepe, J.L. Rivail, B.Robinet, E.Soulié and G.Vergoten - had therefore no great difficulty in bringing together a brilliant group of surveys and contributions. They covered all the ground from general methods and theory to computation techniques. The heart of the conference was of course filled with specific modelling research: catalysis and adsorption, classical methods as applied to model biosystems and pharmacology, quantum models, spectra, kinetic models. In its biophysical aspects molecular modelling had as expected a choice place, dealing with methods, proteins and polypeptides, DNAs, joint modelling-NMR studies.

We had thought equally necessary to assess fields at the borderline of our central theme but essential to our command of computing advances: hence our choice of surveys on the numerical basis of optimization and the "molecular fallout" of research on artificial intelligence.

And since technical developments are crucial we held a panel discussion on the evolution of computer architecture in which IBM, Cray, FPS and Convex took part. The overall presentation of the subject, along with the remarks of great hardware conceivers, make a brief but rewarding chapter.

As always this conference is the work of those who brought a wealth of novel results along with a series of brilliant surveys. In this respect we owe special thanks to several colleagues whom illness or circumstances beyond their control prevented from attending and who were kind enough to send in their contributions: W.F.van Gunsteren, I.Vaisman, T.Solmajer and O.Iordache. We are also pleasantly indebted to many attendees, since we seldom had such lively and profitable discussions. Our thanks to all those who made this symposium a scientific success.

This meeting was also a celebration, that of the role played by Alberte and Bernard Pullman in the developments of Quantum Chemistry and particularly of its applications to Biophysics. They were pioneers on this wonderfully fertile ground, and science owes them research of the highest order along with the formation of many pupils who follow brilliantly on the tracks they have opened. We had great pleasure in building this conference as an homage to the Pullmans, an homage completed by our Society's Council, which elected them as Honorary members.

Finally this 44th International meeting of Physical Chemistry received help and financial support from many organizations and firms to whom we owe much. First of all the University of Nancy I and Nancy City Council, who welcomed us along with the Lorraine Regional Council, Nancy Urban District Authority and the Sciences of Matter Department of the University.

We also received help and support from C.N.R.S., secteur Chimie, Commissariat à l'Energie Atomique (DLPC), Direction des Recherches, Etudes et Techniques, European Office of Aerospace Research and Development (USAF), Hôtels Concorde, I.B.M. France, Ministère de l'Education Nationale (Direction des relations internationales), Ministère de la Recherche et de la Technologie, U.S.Army European Research Office, U.S.Office of Naval Research, London branch. We are grateful for this support, which allowed, among others, to ensure the attendance of young research people and colleagues from countries where research budgets are limited.

Clément Troyanowsky
Honorary Secretary
S.F.C. / Division de Chimie physique

AVANT-PROPOS

Les progrès du graphique moléculaire ont entraîné des avancées très rapides de la modélisation moléculaire, qu'il s'agisse des structures ou des propriétés. La proposition de Jean-Louis Rivail de consacrer notre réunion internationale aux progrès de la modélisation fut donc adoptée pratiquement sans discussion, et le déroulement de cette rencontre a bien montré que nous avions fait le bon choix.

Qu'on me permette ce qui est presque un retour aux origines. Je garde le vif souvenir, étudiant, d'avoir suivi avec passion un colloque tenu à Paris en 1948 sur la liaison chimique, où Mulliken, Pauling, Raman, Louis de Broglie, Ketelaar, Cabannes, Coulson, et les brillants jeunes de la chimie théorique, échangèrent des vues qui - heureusement - ne convergeaient pas toujours. Ce colloque est un peu entré dans la légende car beaucoup d'expérimentateurs, à cette époque, ne voyaient dans la chimie théorique qu'un jeu intellectuel d'intérêt incertain: "On y croira quand les théoriciens seront capables de prédire quelque chose, et pas seulement de démontrer ce qui est déjà connu." Les théoriciens ont depuis fait leurs preuves.

Ils y ont été fabuleusement aidés par les performances chaque jour améliorées des calculatrices. Jeune enseignant j'ai le souvenir d'avoir vu, chez un des grands constructeurs mondiaux, une calculatrice dont l'ensemble occupait environ 100 mètres carrés, et qui calculait "si vite" que ses innombrables erreurs étaient sans importance. Il suffisait de faire chaque calcul une centaine de fois, et de dresser la courbe de Gauss correspondante: le sommet était la valeur cherchée! Ce genre de souvenir fait rêver.

A la même époque l'étude des structures cristallines était un travail de très longue haleine, aujourd'hui il faut bien choisir, pour son intérêt, la structure que l'on établira rapidement. En regard le domaine du "calcul chimique" reste partagé: certains travaux sur la réactivité moléculaire relèvent encore d'un "broyage numérique" laborieux. Dans d'autres domaines, et la modélisation en offre des exemples multiples, des logiciels subtils permettent d'afficher assez simplement les structures et les comportements de molécules complexes.

Aussi notre Comité scientifique - J.Y.Lallemand, R.Lavery, J.P.Mornon, G.Pepe, J.L.Rivail, B.Robinet, E.Soulié et G.Vergoten - n'eut-il guère de difficulté à réunir un brillant ensemble de conférences et de contributions. Elles couvrirent le territoire allant des méthodes générales et de la théorie aux techniques de calcul. Le coeur de la rencontre fut évidemment étoffé par les travaux spécifiques de modélisation: catalyse et adsorption, méthodes classiques de modélisation et leurs applications aux biosystèmes modèles et à la pharmacologie, les modèles quantiques, les spectres, les modèles cinétiques. La modélisation moléculaire dans ses aspects biophysiques: méthodes, protéines et polypeptides, ADN, études conjointes par modélisation et RMN, occupa comme on le prévoyait une place de choix.

Il nous avait paru également nécessaire de faire le point sur des sujets marginaux par rapport à notre thème central, mais essentiels dans tous les développements de l'informatique: on trouvera donc ici une belle introduction aux problèmes de l'optimisation numérique, ainsi que la présentation des retombées "moléculaires" des travaux sur l'intelligence artificielle.

Enfin, vu l'importance cruciale des développements techniques, nous avons tenu une table ronde sur l'évolution de l'architecture des ordinateurs, à laquelle plusieurs des plus grands constructeurs mondiaux ont pris part. La présentation d'ensemble du sujet, comme les remarques des concepteurs de "hardware", forment un chapitre bref mais certainement enrichissant.

Cette réunion, comme toujours, est l'oeuvre de ceux qui y ont apporté une remarquable variété de résultats nouveaux et de vues d'ensemble de grande qualité. A ce sujet nous devons des remerciements particuliers à plusieurs collègues que la maladie ou des circonstances échappant à leur contrôle ont empêché d'être présents mais qui ont tenu à nous fournir leurs contributions: W.F. van Gunsteren, I. Vaisman, T. Solmayer et O. Joradache. Notre colloque a également une agréable dette à l'égard des participants. Nous n'avons pas eu souvent des discussions aussi animées et profitables. Que tous ceux qui ont contribué à la réussite scientifique de ce colloque soient ici remerciés.

Ce colloque était aussi une célébration, celle du rôle joué par Alberte et Bernard Pullman dans les développements de la Chimie quantique et particulièrement de ses applications à la Biophysique. Ils ont été parmi les pionniers de ce terrain si merveilleusement fertile, et la science leur doit des travaux de premier ordre, tout autant que la formation de nombreux disciples qui suivent brillamment les chemins qu'ils ont ouverts. Nous avons été heureux de concevoir cette réunion comme un hommage à leur oeuvre, hommage complété par le Conseil de notre Société, dont les Pullman sont aujourd'hui membres d'honneur.

Enfin cette 44e réunion internationale de Chimie physique a bénéficié de l'appui matériel et financier de nombreux organismes et entreprises auxquels va notre reconnaissance: en premier lieu l'Université de Nancy I et la Municipalité de Nancy pour leur accueil, et les autorités locales et régionales pour leur contribution à notre équilibre budgétaire: Région Lorraine, District urbain de Nancy, UER des Sciences de la matière de l'Université.

Nous avons également reçu les encouragements et l'appui des: C.N.R.S., secteur Chimie, Commissariat à l'Energie Atomique (DLPC), Direction des Recherches, Etudes et Techniques, European Office of Aerospace Research and Development, USAF, F.P.S. Computing, Hôtels Concorde, I.B.M. France, Ministère de l'Education Nationale (Direction des relations internationales), Ministère de la Recherche et de la Technologie, U.S. Army European Research Office, U.S. Office of Naval Research, London branch. Que tous soient assurés de notre gratitude: ces soutiens nous ont permis, entre autres, d'apporter une aide à de jeunes chercheurs et à des collègues de pays dont les budgets recherche sont modestes.

Clément Troyanowsky
Secrétaire général
S.F.C. / Division de Chimie physique

ACKNOWLEDGMENTS

The forty-fourth international meeting of Physical Chemistry has been held at the University of Nancy I, Unité d'Enseignement et de Recherches sur les Sciences de la Matière. We owe it very much for its hospitality as well as for the help we received from many colleagues of the Laboratoire de chimie théorique.

We also received help and support from the following organizations and firms

- Nancy City Council
- Lorraine Regional Council
- Nancy Urban District Authority
- C.N.R.S., secteur Chimie
- Commissariat à l'Energie Atomique, Département des Lasers et de Physicochimie
- Direction des Recherches, Etudes et Techniques
- European Office of Aerospace Research and Development (USAF)
- Hotels Concorde
- I.B.M. France
- Ministère de l'Education Nationale, Direction des Relations Internationales
- Ministère de la Recherche et de la Technologie
- U.S. Army European Research Office
- U.S. Office of Naval Research, London branch

This support is gratefully acknowledged.

LIST OF PARTICIPANTS

- ANDRE J.C./ ENSIC-DCPR/ 1, rue Grandville/ BP 451/ 54001 NANCY Cedex
- ANGYAN J./ Spectrochimie moléculaire/ Univ. P. et M. Curie/ 4, Place Jussieu/ 75252 PARIS Cedex 05
- ARMBRUSTER A.M. Mme/ IBM-France/ Agence Chimie Pétrole/ Tour Septentrion/ La Défense 4 cedex 9/ 92081 PARIS LA DEFENSE
- AUBRY A./ Minéralogie cristallographie/ Univ. Nancy I/ B.P. 239/ 54506 VANDOEUVRE LES NANCY
- AUFFINGER P./ Modélisation moléculaire et de RMN/ 4, rue Blaise Pascal/ 67000 STRASBOURG
- BAYARD M./ Lab. de Chimie Industrielle/ UCBL/Bât.305/ 43,Bd du 11 Novembre 1918/ 69622 VILLEURBANNE Cedex
- BERGEL A./ Lab.génie chimique UA 192/ Univ. Paul Sabatier/Lab. d'électrochimie/ 118, route de Narbonne/ 31062 TOULOUSE Cedex
- BERGES J. Mme/ Dynamique des Int. Moléculaires/ Tour 22/ Couloir 22-23/1er étage/ 4, place Jussieu/ 75005 PARIS
- BIZOT H./ LPCM.INRA/ B.P.526/ 44026 NANTES Cedex 03
- BLEGER A. Mlle/ BIOSTRUCTURES/ 8, rue Gustave Adolphe Hirn/ 67000 STRASBOURG
- BONNAMOUR F./ CRAY Research France/ 7, rue de Tilsitt/ 75017 PARIS
- BOUDON S./ Institut de Chimie/ Lab. RM3/UA422/ 4, rue Blaise Pascal/ 67000 STRASBOURG
- BOUQUANT J./ Chimie organique physique/ UFR Sciences/ B.P.347/ 51062 REIMS cedex
- BOURAOUI A./ ESIPSOI./UA 126/CNRS/ Fac.Sciences St Jérôme/ Av. Normandie Niemen/ 13397 MARSEILLE Cedex13
- BRUNEAU P./ ICI Pharma/ Chemin de Vrilly/ Z.I. la Pompelle/ BP 401/ 51064 REIMS Cedex
- BRUNETIERE A./ Rhone Poulenc Agrochimie/ CRLD/ 14 à 20 rue P. Baizet/ 69009 LYON
- BUCKINGHAM A.D./ Cambridge Univ./ Univ. Chemical Lab/ Lensfield Rd/ CAMBRIDGE CB2 1EW/(G.B.)
- CAILLET J./Dynamique des Int. Moléculaires/ Tour 22/couloir 22-23/1er étage/ 4, place Jussieu/ 75005 PARIS
- CARLES P./ St Procida/ B.P.1/ 13367 MARSEILLE Cedex 11
- CASTILLO S. Mme/ Lab. recherche sur l'énergie/ Univ. Paul Sabatier/ 118, route de Narbonne/ 31062 TOULOUSE Cedex

XVIII

CASTRO B./ SANOFI Chimie/ 17, rue des Fossés St Jacques/ 75005 PARIS

CENSE J.M./ E.N.S.C.P./ 11, rue P. et M. Curie/ 75231 PARIS Cedex 05

CHANTELOUBE M./ BIOSYM Technologies Inc./ 15, av. Victor Hugo/ 75116 PARIS

CHERMETTE H./ IPN/ Lab. Chimie nucléaire/ 43, Bd du 11 Novembre 1918/ 69622 VILLEURBANNE Cedex

CHOJNACKI H./ Inst. of Organic and Physical Chemistry/ Wyb. Wyspianskiego 27, 1-4/ 50-370 WROCLAW/ (Pologne)

CHOMILIER J./ Biophysique/ Muséum national d'histoire Naturelle/ 43, rue Cuvier/ 75231 PARIS Cedex 05

C.N.R.S./I.N.I.S.T./ 26, rue Boyer 75020 PARIS

COLETTI PREVIERO M.A. Mme/ INSERM/U 58/ rue de Navacelle/ 34000 MONTPELLIER

COLONNA F./ D.I.M./ Tour 22/1er étage/ 4, place Jussieu/ 75252 PARIS

COULOMBEAU C./ LEDSS VI/ Univ. J. Fourier/ BP 53 F/ 38041 GRENOBLE Cedex

CRASPAIL N. Mlle/ CRMC2-CNRS/ Campus Luminy Case 913/ 13288 MARSEILLE Cedex

CREUZET S. Mlle/ CEA/INRA/ CEN Saclay/ SBPH Bât. 532/ 91191 GIF SUR YVETTE

DASSONVILLE R./ CRMC2-CNRS/ 70, route Léon Lachamp/ Case 913/ Campus de Luminy/ 13288 MARSEILLE Cedex 2

DAUTANT A./ Cristallographie/ Univ. Bordeaux I/ 33405 TALENCE Cedex

DECORET C./ Lab. de Chimie Industrielle/UCBL/ UA 805-CNRS/ Bât.305/ 43, Bd du 11 Novembre 1918/ 69622 VILLEURBANNE Cedex

DEHARENG D. Mme/ Univ. de Liège/ Lab. de Microbiologie/ Inst. de Chimie/Bât. B6/ SART TILMAN/ 4000 LIEGE(Belgique)

DELANCE E./ ITODYS/ 1, rue Guy de La Brosse/ 75005 PARIS

DEMANGE P./ Chimie physique macromoléculaire/ ENSIC-INPL/ BP 451/ 1, rue Grandville/ 54001 NANCY Cedex

DERREUMAUX P./ Faculté de Pharmacie/ INSERM U279/ 59045 LILLE Cedex

DESMARQUETS M./ BIOSTRUCTURES/ 8, rue Gustave Adolphe Hirn/ 67000 STRASBOURG

DEVILLERS J./I.M.R.C.P./ Univ. P. Sabatier/ 118, route de Narbonne/ 31400 TOULOUSE Cedex et Lab. chimie de coordination, 205, route de Narbonne/ 31077 TOULOUSE Cedex

DIVE G./ Lab. microbiologie/ Univ. de Liège/ Institut de Chimie/Bât B6/ SART TILMAN/ 4000 LIEGE(Belgique)

DURUP J./ Physique quantique/ Univ. Paul Sabatier/ 118, route de Narbonne/ 31062 TOULOUSE Cedex

DYMEK Chester J. Jr Lt-Col./E.O.A.R.D./223 Old Marylebone Rd/LONDON NW1 5TH (G.B.)

ETCHEBEST C. Mlle/ IBPC/ 13, rue P. et M. Curie/ 75005 PARIS

EGMOND M.R./ Unilever Research Lab./ PO BOX 114/ 3130 AC VLAARDINGEN/(Pays Bas)

FAIEZ R./ Chimie théorique/ Univ. Nancy I/ BP 239/ 54506 VANDOEUVRE LES NANCY

FATHALLAH M./ESIPSOI UA 126/ Fac. des Sciences St Jérôme/ Av. Escadrille Normandie Niemen/ 13397 MARSEILLE Cedex 13

FAUPEL P./ Univ. Gh. Duisburg/ FB6, Theoretische Chemie/ Lotharstrasse 1/ 4100 DUISBURG 1/ RFA

FAVROT J./ Structure et réactivité de molécules phosphorées/ Univ. P. Sabatier/ 118, route de Narbonne/ 31062 TOULOUSE Cedex

FOURCOT F./ ESPCI/ Lab. Chimie analytique/ 10, rue Vauquelin 75231 PARIS Cedex 05

FRITSCH V. Mlle/ I.B.M.C./ 15, rue R. Descartes/ 67084 STRASBOURG

GALLEYRAND G./ IBM France/ 67, quai de la Rapée/ 75012 PARIS

GALLO R./ ESIPSOI/ Fac. Sciences St Jérôme/ Av Normandie-Niemen/ 13397 MARSEILLE Cedex 13 F

GAYDOU E./ Phytochimie/ Fac. des Sciences St Jérôme/ Av. Escadrille Normandie Niemen/ 13397 MARSEILLE Cedex 13

GELIN B./ Polygen(Europe)Ltd/ 25, rue du Pont des Halles/ 94566 RUNGIS Cedex

GENEST D./ CBM/CNRS/ 1A, av. de la Recherche Scientifique/ 45071 ORLEANS Cedex 2

GENEST M. Mme/ CBM/ C.N.R.S./1A av. de la Recherche Scientifique/ 45071 ORLEANS Cedex 2

GHOMI M./ Lab. de Spectroscopie Biomoléculaire/ UFR Biomédicale de Bobigny/ Univ. Paris XIII/ 74, rue Marcel Cachin/ 93012 BOBIGNY Cedex

GILQUIN B./ CISI-CEA/ Service de Biochimie/ Depart. Biologie/ 91191 GIF SUR YVETTE

GINGOLD M./ Biophysique/ Depart. de Biologie/ CEN SACLAY/ 91191 GIF SUR YVETTE Cedex

GIRE P./ CONVEX/ Parc d'Activités du Pas du Lac/ Immeuble le Daguerre/ 9, av. Ampère/ 78180 MONTIGNY LE BRETONNEUX

GRAND A./ DRF/Lab. de Chimie/ CENG/85X/ 38401 GRENOBLE Cedex

GRESH N./ Biochimie théorique/ I.B.P.C./ 13, rue P. et M. Curie/ 75005 PARIS

GUILLOT B./ Physique théorique des liquides/ Tour 16/ 4, place Jussieu/ 75252 PARIS Cedex 05

HAAS V. Mlle/ Univ. GH Duisburg/ FB6/ Theoretische Chemie/Lotharstrasse 1/ 4100 DUISBURG 1 (RFA)

HARTMANN B. Mme/ IBPC/ 13, rue P. et M. Curie/ 75005 PARIS

XX

HATON M.C. Mme/ CRIN, Univ. de Nancy I/B.P. 239/54506 VANDOEUVRE LES NANCY

HENRIET C./ CRAY RESEARCH FRANCE/ 7, rue de Tilsitt/ 75017 PARIS

HERSCOVICI A./ IBM France/ Tour Descartes/ LA DEFENSE 5/ 2, av. Gambetta/
92400 COURBEVOIE

HOGGAN P./ Lab. Chimie théorique/ Univ. Nancy I/ B.P. 239/ 54506 VANDOEUVRE
LES NANCY

HUIGE C. J.M./ Gorlaeus Lab./ State Univ. Leiden/ PO BOX 9502/ 2300 RA
LEIDEN/(Pays-Bas)

HUNGARO J.B./ Chimie physique/ Univ. Paris VII/ Tour 53-54 5ème étage/ 2,
place Jussieu 75005 PARIS

IORDACHE O./ Polytechnical Inst. of Bucharest/ Depart. of Chemical Eng./
1 Polizu/ R-78126 BUCHAREST 12/(Roumanie)

JOFFRE J./ Chimie organique physique et Cinétique chimique appliquées/ ENSC/
8, rue de l'Ecole Normale/ 34075 MONTPELLIER Cedex

KARPLUS M./ Depart. of Chemistry/ Harvard Univ./ 12, Oxford Street/
CAMBRIDGE, Massachusetts 02138 (U.S.A.)

KASSAB E./ D.I.M./ Tour 22/1er étage/ Univ. P. et M. Curie/ 4, place Jussieu/
75252 PARIS Cedex

KOCHANSKI E. Mme/ ER 139/ Chimie théorique/ Inst. de Chimie/ 1, rue Blaise
Pascal/ B.P. 296/R8/ 67008 STRASBOURG

KOZELKA J./ Univ. René Descartes/ Chimie et Biochimie Pharmacologiques et
toxicologiques/ 45, rue des Saints Pères/ 75270 PARIS Cedex 06

KWIATKOWSKI J.S./ N. Copernicus Univ./ Inst. of Physics/ 87-100 TORUN
(Pologne)

LACROIX-DORE M.D.Mme/ IBM FRANCE/ 67, quai de la Rapée/ 75012 PARIS

LAGANT P./ Fac. Pharmacie Lille/ INSERM U 279/ 3, rue du Professeur Laguesse/
59045 LILLE Cedex

LAHANA M./ Aquitaine Systèmes/ Tour Elf/ 2, place de la Coupole/ 92078 PARIS
LA DEFENSE

LAMBERT C. Mme/ I.C.I. PHARMA/ Centre de Recherches/ Chemin de Vrilly/ Z.I.
la Pompelle/ B.P. 401/ 51064 REIMS Cedex

LANGLET G/ CEA/IRDI/DLPC/SCR/ CEN SACLAY/ 91191 GIF SUR YVETTE

LANGLET J. Mme/ Dynamique des Int. Moléculaires/ Tour 22/ Couloir 22-23/
4, place Jussieu/ 75005 PARIS

LAPIANTE S./ ICSN-CNRS/ 91190 GIF SUR YVETTE

LAUPRETRE F. Mme/ Physico-chimie structurale et macromoléculaire/ ESPCI/
10, rue Vauquelin/ 75231 PARIS Cedex 05

LAVERY R./ I.B.P.C./ 13, rue P. et M. Curie/ 75005 PARIS

LECLERC J.P./ Centre de recherches de Vitry/ B.P. 14/ 13, quai Jules Guesde/
94403 VITRY S/ Seine Cedex

LECLERCQ J.M./ Lab. dynamique Int. moléculaires/ UPR A0271 CNRS/ Univ.P.
et M. Curie/ Tour 22/ 4, place Jussieu/ 75252 PARIS Cedex 05

LEGOAS M./ D.C.S.O./ Ecole Polytechnique/ 91128 PALAISEAU

LEMARECHAL C./ INRIA/ B.P.105/ 78153 LE CHESNAY

LLUCH J./ Depart. Quimica/ Univ. Autonoma de Barcelona/ 08193 BELLATERRA
(Barcelona)/(Espagne)

LOOS M./ Lab. Chimie théorique/ Univ. Nancy I/ B.P. 239/ 54506 VANDOEUVRE
LES NANCY

MAISSIAT C./ Société Biomodèle/ CEI/ 15, rue Carnot/ BP 175/ 86004 POITIERS
Cedex

MAJOUBE M./ CEA-DLPC/ CEN SACLAY/ 91191 GIF SUR YVETTE

MALLIAVIN T. Mlle/ DESO/ Ecole Polytechnique/91128 PALAISEAU Cedex

MARK F./ Max Planck Inst. fur Strahlenchemie/ Stiftstr 34-36/ D 4330
MULHEIM/(RFA)

MAROUN R./ Inst. Biologique/ Pharmacochimie moléculaire/ 4 av. de l'Obser-
vatoire/ 75006 PARIS

MASSEY K./ FPS COMPUTING/ 21, rue des Cévennes/ Silic 523/ 94633 RUNGIS

MATHIEUX J.P./ CRAY RESEARCH FRANCE/ 7, rue de Tilsitt 75017 PARIS

MATHIS H./ Chimie théorique/ Univ. Nancy I/ B.P. 239/ 54506 VANDOEUVRE LES
NANCY

MATTALIA J.M./ Chimie inorganique/ Fac. des Sciences St Jérôme/ 13397 MAR-
SEILLE Cedex 13

MAUREL J.L./ Procida/ Groupe Roussel-Uclaf/ ST MARCEL/ 13011 MARSEILLE

MAZEAU K./ CERMAV,CNRS/ B.P. 53x/ 38041 GRENOBLE Cedex

MEHANI S./ BP RESEARCH Center/ Chertsey Road/ Sunbury on Thames/ MIDDLESEX,
LDN, TW13 7 LN/UK

MIKOU A. Mlle/ ICSN/CNRS/Lab. RMN/ rue de la Terrasse/ 91190 GIF SUR YVETTE

MILLOT C./ Lab. Chimie théorique/ Univ. Nancy I/ B.P. 239/ 54506 VANDOEUVRE
LES NANCY

MILON A./ URA 31-CNRS/ Centre de Neurochimie/ 5, rue Blaise Pascal/ 67084
STRASBOURG

MONQUE R./ INTEVEP SA/ Apdo 76343 CARACAS 1070 A/(Venezuela)

MONTEIL P./ CONVEX/ Parc d'Activités du Pas du Lac/ Immeuble le Daguerre/
9, av. Ampère/ 78180 MONTIGNY LE BRETONNEUX

XXII

MOREAU G./ ROUSSEL UCLAF/ 102, route de Noisy/ 93230 ROMAINVILLE

MORGANTINI P.Y./ Depart. chimie physique/ Univ. de Genève/ 30, quai E. Ansermet/ 1211 GENEVE 4 (Suisse)

MORIZE I. Mme/ Rhone Poulenc Santé/ Modélisation moléculaire/ 13, rue Jules Guesde/ 94403 VITRY sur Seine Cedex

MOY L./ Aquitaine Systemes/ Tour Elf/ 2, place de la Coupole/ 92078 PARIS LA DEFENSE

MULLER N. Mme/ Inst. Henri Beaufour(IHB)/ 17, av. Descartes/ 92350 LE PLESSIS ROBINSON

NICKLAS K./ Inst. fur Physikalisc' Chemie/ TH Darmstadt/ Petersenstr. 20/ 6100 DARMSTADT/(RFA)

OHLENBUSCH H./ Inst. de Physique biologique/ 4, rue Kirschleger 67085 STRASBOURG Cedex

PANTEL G./ Rhone Poulenc Santé/ Centre de recherches de Vitry/ BP 14/ 13, quai Jules Guesde/ 94403 VITRY sur SEINE Cedex

PENTENERO A./ UER des Sciences de la matière/ Univ. de Nancy I/ B.P. 239/ 54506 VANDOEUVRE LES NANCY

PERAHIA D./Lab. d'Enzymologie physico chimique et moléculaire/ Univ. Paris Sud/ Bât. 430/91405 ORSAY

PEPE G./ CRMC2-CNRS/ Campus Luminy/ Case 913/ 13288 MARSEILLE Cedex

PESQUER M./ Physicochimie théorique/ CNRS-UA 503/ Univ. Bordeaux I/ 33405 TALENCE Cedex

PETITTT B.M./ Chemistry department/ Univ. of Houston/ 4800 CALHOUN/ HOUSTON TEXAS 77204-5641/(USA)

PIRIOU J.M./ Biochimie-Enzymologie/ Inst. Gustave Roussy/ 39, rue Camille Desmoulins/ 94805 VILLEJUIF Cedex

PLOCKYN M./ Jobin Yvon/16-18 rue du Canal/ 91160 LONJUMEAU

POLTEV V./ Inst. of Biological Physics/ USSR Academy of Sciences/ Pushchino, Moscow Region/ 142292 USSR

PONCIN M./ IBPC/ 13, rue P. et M. Curie/ 75005 PARIS

POTHIER J./ Biologie Enzymologie/ inst. Gustave Roussy/ 39, rue Camille Desmoulins/ 94805 VILLEJUIF Cedex

PREMILAT S./ Biophysique moléculaire/ B.P. 239/ 54506 VANDOEUVRE LES NANCY

PULLMAN A. Mme/ Inst. de Biologie Physico Chimique/ 13, rue P. et M. Curie/ 75005 PARIS

PULLMAN B./ Inst. de Biologie Physico chimique/ 13, rue P. et M. Curie/ 75005 PARIS

RAHMOUNI A./ Chimie théorique/ Inst. de Chimie/ 1, rue Blaise Pascal/ 6700 STRASBOURG

REPELIN Y. Mme/ Ecole Centrale/ Grande voie des vignes/ 92290 CHATENAY MALABRY

REYNAUD J.C./ CONVEX/ Parc d'Activités du Pas du Lac/ Immeuble le Daguerre/
9, rue Ampère/ 78180 MONTIGNY LE BRETONNEUX

RINALDI D./ Chimie théorique/ Univ. Nancy I/ B.P. 239/ 54506 VANDOEUVRE
LES NANCY

RIVAIL J.L./ Lab. Chimie théorique/ Univ. Nancy I/ B.P. 239/ 54506 VANDOEUVRE
LES NANCY

ROBERT B./ CENS/SBPH/CEN SACLAY/ 91191 GIF SUR YVETTE

ROZOT M./ Lab. Chimie théorique/ Univ. Nancy I/ B.P. 239/ 54506 VANDOEUVRE
LES NANCY

RUBIN C. Mlle/ Inst. de Chimie/ 1, rue Blaise Pascal/ 6700 STRASBOURG

RUIZ M./ Chimie théorique/ Univ. Nancy I/ B.P. 239/ 54506 VANDOEUVRE LES
NANCY

RULLMANN J.A.C./ Department of Organic Chemistry/ Univ. of Utrecht/ Padu-
alaan 8/ 3584 CH UTRECHT/ (Pays Bas)

SAARMETS A./ Synthelabo Recherche/ 31, av. P. Vaillant Couturier/ 92200
BAGNEUX

SCHLENKRICH M./ Inst. fur Physikalische Chemie/TH. Darmstadt BRD/ Peter-
senstr. 20/ 6100 DARMSTADT/ (RFA)

SCHWAAB F./CRIN/ B.P. 239/ Univ. Nancy I/ 54506 VANDOEUVRE LES NANCY

SHUSTOROVICH E./ Corporate Research Lab./ Eastman Kodak Company/ Corl
B-81/ ROCHESTER, NY 14650-02001/(U.S.A.)

SIRI D./ CRMC2-CNRS/Campus Luminy/ Case 913/ 13288 MARSEILLE Cedex 9

SMITH J./ CISI-CEA/ Service de Biophysique/ Depart. de Biologie/ 91191
GIF SUR YVETTE

SOLMAJER T./ Boris Kidric Chemistry Inst./ Hajdrihova 19/ 6100 LJUBLJANA
SLOVENIA/ (Yougoslavie)

SOULIE E./ IRDI/DESICP/DIPC/SCM/CEN SACLAY/91191 GIF SUR YVETTE

SOUMPASIS D.M./ Max Planck Inst. fur Bioph. Chemistry/ P.O. Box 2841/
3400 GOTTINGEN/(R.F.A.) et Los Alamos Nat. Lab./ LOS ALAMOS 87545/ (USA)

STAWARZ B./ Centre de Biophysique Moléculaire/ 1A, av. de la Recherche
Scientifique/ 45071 ORLEANS Cedex 2

STONE A.J./ Univ. Chemical Lab./ Lensfield Road/ CAMBRIDGE CB2, 1 EW (GB)

SUN Jian-Sheng/ Biophysique, Inserm 201/ CNRS UA 481/ Museum National
d'Histoire Naturelle/ 43, rue Cuvier/ 75031 PARIS

SURCOUF E.Mme/ Rhone Poulenc Santé/ Centre de recherches de Vitry/ BP
14/ 13, quai Jules Guesde/ 94403 VITRY sur Seine Cedex

TAYLOR J./ BIOSYM Technologies Inc./ 15, av. Victor Hugo/ 75116 PARIS

XXIV

TORRENS F./ Departament de Quimica Fisica/ Univ. de Valencia/ Dr. Moliner
50/ 46100 BURJASSOT (Espagne)

TRINQUIER G./ Physique quantique/ Univ. Paul Sabatier/ 118, route de Narbonne/
31062 TOULOUSE Cedex

TROYANOWSKY C./ Division de Chimie physique/ 10, rue Vauquelin/ 75005
PARIS

URBAIN D./ FPS COMPUTING/ 21, rue des Cévennes/ Silic 523/ 94633 RUNGIS

VAISMAN I.I./ Ins. of non Aqueous Solutions/ Chemistry of the USSR/ Academy
of Sciences/ Akademicheskaya Street I/ 153045 IVANOV/ (U.S.S.R.)

VAN GUNSTEREN W.F./ Depart. of Physical Chemistry/ Univ. of Groningen/
Nyenborgh 16/ 9747AG GRONINGEN/ (Pays Bas)

VERGELATI C./ Rhone Poulenc Recherches/ Centre de recherches des Carrières/
85, av. des Frères Perret/ B.P. 62/ 69192 SAINT FONS Cedex

VERGOTEN G./ Fac. de Pharmacie Lille/ INSERM U279/ 3, rue du Professeur
Laguesse/ 59045 LILLE Cedex

VIGNE F. Mme/ I.R.C./ 2, av. Albert Einstein/ 69626 VILLEURBANNE Cedex

VIRIOT M.L. Mme/ DCPR/ ENSIC-INPL/ 1, rue de Grandville/ 54000 NANCY

VOISIN C./ Chimie théorique/ Univ. de Nancy I/ B.P. 239/ 54506 VANDOEUVRE
LES NANCY

VOLTZ R./ C.R.N./ Univ. Louis Pasteur/ B.P. 20/ 67037 STRASBOURG Cedex

WARSHEL A./ U. Southern California, Chemistry/ U. Park/ LOS ANGELES, CA
90089-1061/ (USA)

WIPFF G./ RMN et Modélisation moléculaire/ Inst. de Chimie/ 1, rue Blaise
Pascal/ 67000 STRASBOURG

WODAK S. Mme/ Univ. Libre de Bruxelles/ UCMB/ CP 160/ Av. Paul Heger P2/
B 1050 BRUXELLES/ (Belgique)

WURFLINGER A./ Lehrstuhl für Physikalische Chemie II/ Ruhr Univ. Bochum/
Universitätsstrasse 160/ 4630 BOCHUM I/ (RFA)

ZAKRZEWSKA K. Mme/ I.B.P.C./ 13, rue P. et M. Curie/ 75005 PARIS

MOLECULAR MODELLING, with or without QUANTUM CHEMISTRY ?

Bernard PULLMAN, Institut de Biologie Physico-Chimique, 13, rue Pierre et Marie Curie, 75005 Paris (France)

"A good model is worth its weight in gold".

Francis Crick in "What Mad Pursuit. A personal view of Scientific Discovery" Basic Books Inc. Publishers. New York 1988 p. 86.

Coming from Francis Crick and obviously inspired by his personal experience, which is that of one of the greatest discovery ever made in biology revealing one of the essential secrets of the mechanism of life on earth, the above quotation represents for all those engaged or willing to engage in molecular modelling an exceptional encouragement. It reminds us also of a fact which strangely or amusingly seems to be ignored by a large number of young biologists and even forgotten by some of older ones, namely that at the moment of their discovery of the double-helical structure of DNA, Watson and Crick have never carried out any experimental work, neither by X-ray crystallography nor by any other means, on this biopolymer. Even better, to quote Crick (ref. 1, p. 67) : "One of the oddities of the whole episode is that neither Jim nor I were officially working on DNA at all. I was trying to write a thesis on the X-ray diffraction of polypeptides and proteins, while Jim has ostensibly come to Cambridge to help John Kendrew crystallize myoglobin" and elsewhere (ref. 1, p. 65) : "Following Pauling's example, we believed the way to solve the structure was to build models. The London workers followed a more painstaking approach". The London workers to which Crick refers are Maurice Wilkins and Rosalind Franklin and the painstaking approach, the constant refinement of the X-ray diffraction patterns.

In the last quotation Crick refers to Pauling's example and what he has in mind is Pauling's then very fresh discovery (in 1951) of the α -helical structure of polypeptides, another milestone in the history of modern molecular biology. In relation to the title of my lecture this may be the first good case to look at more deeply.

One may recall that in the midst of the century two important Laboratories were striving (competing?) for the establishment of the structure of polypeptides and proteins by building models which would permit to interpret correctly the then available X-ray diagrams of keratin in particular. These two groups were those of Pauling and collaborators in California, and of Bragg, Perutz and Kendrew in Cambridge, England. As is well known it is the first of these groups which won the race. What was the essential advantage which gave Pauling the victory ? A very simple piece of

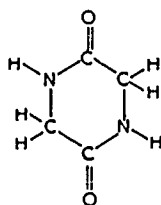
knowledge, which he possessed, owing at least in large part to his quantum chemical approach, namely the planarity of the peptide bond which results from the partial double bond character of its C-N link, resulting itself from the "resonance" (conjugation) of the π -electrons of the C=O double bond with the lone pair electrons of the N atom. This knowledge simplified greatly the model building and played an important role in leading to the α -helix. This is perhaps the first example of a decisive intervention of a quantum chemical concept in the modelling of a molecular system of important, in this case fundamental, biological significance. The English group lacked this knowledge and was consequently too permissive about the flexibility of the system, which was one of the reasons which prevented it from finding the good answer. The core of the story is, however, still more perplexing, amusing or shocking depending on how you look on it, especially for those of us who are quantum chemists and especially for those who belong to the somewhat older generation. I am quoting again from Crick (ref. 1, p. 58). "Perutz learned that after one of his seminars a local physical chemist had told him that the peptide group ought to be planar. Perutz has even recorded it on his notes, but had done nothing about it. It was not that they had not tried to get good advice, but some of what they had received had been unfortunate. Charles Coulson, a theoretical chemist from Oxford, had told them, in my hearing, that the nitrogen atom might be "pyramidal", which was a highly misleading piece of information". The name of the physical chemist, who gave the good advice is not given but Coulson whose blunder is, truly, unbelievable is committed to scientific damnation. I wonder whether his name would have been quoted if he were from Cambridge instead of being from Oxford. A small consolation to his darkened memory may come from more recent findings which indicate that the planarity of the peptide bond in proteins is not always absolute and that even strong deviations from it may occur at some places in some proteins. As early as 1968 Ramachandran, who earlier masterfully utilized the idea of planar peptide bonds to construct his well known ϕ , ψ Ramachandran plots (for general information see e.g. ref. 2) indicated, in a paper bearing this very little "the need for nonplanar peptide units in polypeptide chains". In fact later developments have shown that very small energy expense is needed for twisting the peptide bond up to 20-30° (see e.g. ref. 2). But these are refinements and Coulson's blunder remains historical.

Coming back to the DNA double-helix, was quantum chemistry also of any (or some) significance in its discovery? The answer is again yes and it resides again in a piece of knowledge based or supported by quantum mechanics which turned out to be decisive for the construction of the correct model. This time this piece of knowledge relates to the correct representation of the dominant tautomeric forms of the nucleic acid bases, in particular in relation to their possible Lactam-Lactim (keto-enol) tautomerism.

The representation prevalent at that time (circum 1950) and currently found then in all textbooks described the oxygen containing bases (G, C and T or U) as existing in the lactim (enol) form. As it is obvious to all of us today this situation precluded the correct representation of the hydrogen-bonded base-pairing scheme (G-C, A-T) and consequently the building of the correct model of DNA. It was against this difficulty that Jim Watson stumbled

for months till the fateful day in which an american visitor at the Cambridge Laboratory Jerry Donohue (this time the name is known) indicated to him that, following his opinion, the bases existed in the Lactam (keto) form. Now, quoting from Jim Watson's celebrated book (ref. 3, p. 192) : "He (Jerry Donohue) admitted that only one crystal structure bore on the problem (preference for the keto form). This was diketopiperazine whose three dimensional configuration had been carefully worked out in Pauling's Laboratory several years before. Moreover he felt sure that the quantum-mechanical arguments which showed why diketopiperazine has the keto form should also hold for guanine and thymine. I was thus urged not to waste more time with my hare-brained scheme". Whatever it be and we shall come to this point back immediately, "after that the (correct double helical) model was almost inevitable" (ref. 1, p. 85) and, indeed, the next day the double helix was born (ref. 3, p. 194).

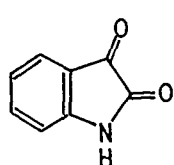
I was intrigued by the "quantum-mechanical arguments" referred to in the above quotation and went to see the original paper (ref. 4). I am sorry to admit that the arguments that I found there while sufficiently suggestive do not appear as completely persuasive. Diketopiperazine as seen in (I), is a cyclic dipeptide which contains two "isolated" peptide bonds, in the sense that they are separated by saturated carbons and are thus not involved in a "general" molecular "resonance" (electronic delocalization), typical of the



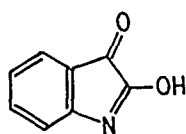
I. Diketopiperazine

"conjugated" electronic systems of the heteroaromatic nucleic acid bases. The "arguments" advanced by Corey refer essentially to the electron delocalization within each peptide bond which determines their planarity and to the overall observed quasi planarity of the molecule. They could not prejudge thus in fact of what may happen for peptide bonds inserted within a larger conjugated system such as that characteristic of the nucleic acid bases. The situation in the later type of compounds, as it could be considered objectively at that time may be inferred from the following paragraph in our book "Les Théories Electroniques de la Chimie Organique" (Masson Eds. Paris, p. 259) published in 1952, thus just at the eve of the discovery of the double helix : "Les exemples classiques de la tautomérisation lactame-lactime (dans les composés conjugués) sont fournis par l'isatine, XXXVIIa et XXXVIIb, la saccharine, XXXVIIIa et XXXVIIIb, le carbostyrile XXXIXa et XXXIXb, etc. La forme lactame est, du point de vue de la somme des énergies des liaisons, plus stable que la forme lactime d'à peu près

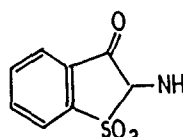
10 kcal/mol. En revanche, la forme lactime a, dans les composés indiqués, une énergie de résonance légèrement plus grande que la forme lactame, et cela grâce au transfert de la double liaison à l'intérieur du cycle. La position exacte de l'équilibre tautomère dans ces corps est toutefois inconnue".



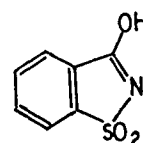
XXXVII a



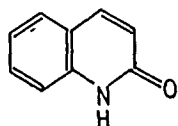
XXXVII b



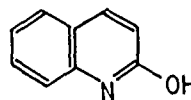
XXXVIII a



XXXVIII b



XXXIX a



XXXIX b

It seems thus that Donohue's "arguments" were in fact partly an inspired guess. His generalization turned out indeed to be correct and played thus a role in shaping the history of biology. [The general problem of the tautomerism of the purine and pyrimidine bases turned out to be nevertheless of importance for some biological problems e.g. mutations. At the time when experimental studies were difficult Quantum Theory played a role in exploring various features of this phenomenon. (For general reviews see refs. 5,6)].

And what about Pauling? Well, remarkable chemist as he was, he was following then a completely erroneous track, proposing a model for DNA composed of three helices with the sugar phosphate backbone in the center and the bases at the periphery. The most astonishing feature of the model, surprising for the genial quantum chemist that he was, and still is, was that the phosphate groups of his model were not ionized. The hydrogens of the phosphates were on the contrary used to form an H-bonding scheme that held the three intertwined chains together. Take these H-bonds away and the structure collapses. As it did. As Watson says (ref. 3, p. 160) "Pauling's nucleic acid was not an acid at all". An astonishing blunder.

Enough about the past. And what about the present? Certainly a useful question at a time when molecular modelling becomes an important instrument of research in chemistry, biology, pharmacology, etc. Is quantum chemistry still of use in this respect or can it be superseded by more classical approaches? You will certainly not be astonished if I defend the

significance of quantum chemistry in this respect and I will illustrate my point on a very topical example.

It concerns DNA again and relates to the search of factors responsible for sequence selectivity in the interaction of groove binding antitumor ligands with this biopolymer.

The most representative and best known ligands of this series, exemplified by netropsin and distamycin A (Fig. 1) show a marked triple specificity for binding to : 1) the minor groove, 2) of AT sequences 3) of B-DNA for a recent review (refs. 7,8).

The current and historically first proposal attributes this specificity essentially to hydrogen bond formation between the peptidic NH groups of the drugs and the O2 atoms of thymine and/or N3 atoms of adenine situated in the minor groove of DNA. Within this proposal the charged end groups of the drugs are considered to be also involved in the interaction probably with the phosphate groups of DNA. The reality of the hydrogen bonds is confirmed by an X-ray study of the crystal structure of a complex between netropsin and the double-helical DNA dodecamers CGCGAATT^{Br}CGCG (ref. 11), (CGCGATATCGCG)₂ (ref. 12) and by NMR studies on the association of netropsin and distamycin A with AT sequences in oligonucleotides (refs. 13,14).

That the situation is, however, more complicated than this simple picture suggests and that, in particular, the precise role of the hydrogen bonds in determining specificity has to be reconsidered becomes evident from the examination of other molecules studied. Thus, in particular the bisquaternary ammonium heterocycle SN 18071 (Fig. 1), which has no hydrogen bonding possibilities, binds also to DNA and shows a similar AT minor groove specificity (refs. 15,16).

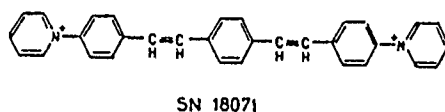
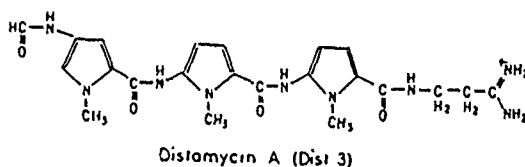
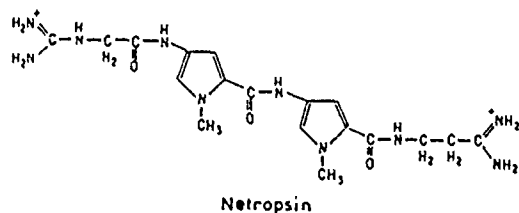


Fig. 1. Typical AT minor groove binding ligands.

An indication that the source of the specificity common to these diverse drugs (and many others) may reside to a large extent in the properties of the minor groove of AT sequences of B-DNA, rather than in special features of the drugs, was suggested first (ref. 17) by the observation that the grooves are the sites of location of the deepest molecular electrostatic potential in DNA, that for AT sequence the deepest potential occurs in their minor groove and that the deepest potentials in DNA are those of the minor groove of AT sequences (refs. 18,19). One could conceive then that provided that this groove could offer an appropriate steric fit to the drugs involved, the origin of specificity could reside in the combination of this fit with a corresponding strong electrostatic interaction.

The electrostatic molecular potential referred to above is a typical quantum-mechanical index of the electronic structure of molecules. It defines (refs. 20,21) the electrostatic (Coulomb) potential created in the neighbouring space by the nuclear charges and the electronic distribution of the system under investigation. For a given wave function with the corresponding electron distribution function $\rho(i)$ the value of such a potential at a given point P in space $V(P)$, is given by :

$$V(P) = \sum_{\alpha} \frac{Z_{\alpha}}{r_{\alpha P}} - \int \frac{\rho(i)}{r_{Pi}} d\tau_i$$

where Z_{α} is the nuclear charge of nucleus α . This quantity has the double advantage of being directly obtainable from the wave function and of being an expression of the global molecular reality, clearly related to what a reactant "feels" upon approaching the substrate.

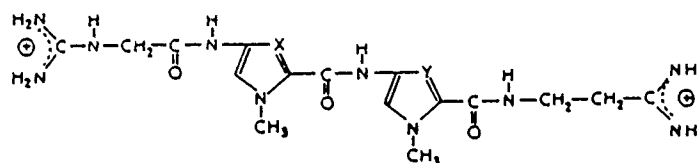
The hypothesis of the decisive significance of this potential for the selective interaction of the minor groove binding ligands with DNA was rapidly confirmed by explicit computations of interaction energies between a number of compounds of the type illustrated in Fig. 1 with model poly(dA).poly(dT) and poly(dG).poly(dC) duplexes in B-DNA conformation. Whatever the approximations used in the computations (free space interaction (ref. 22), interactions in solution (refs. 23,24)) the results invariably show, for all the compounds investigated that, out of the four possibilities of binding namely to the AT minor groove, AT major groove, GC minor groove or GC major groove, the greatest values of the interaction energy are obtained for interaction with the minor groove of the AT sequences. The details of the computations indicate that this preference is favoured predominantly by the electrostatic component of the interaction energy. This demonstrated that whatever the significance of hydrogen bonds for the stability of the complex, the formation of these bonds is not necessary neither for binding nor for the preference for the minor groove of the AT sequences of B-DNA. It seems, in conformity with the original hypothesis, that provided that a steric fit can

be obtained in the minor groove the ligand will be sufficiently stabilized there by the favourable electrostatic potential generated by the AT sequences. When possible, hydrogen bonds between the proton donating sites of the ligand and the proton accepting sites of the macromolecule are, of course, formed and contribute to the energy of binding as indicated by the greater values of that energy in netropsin and distamycin A than in SN 18071. Interestingly for the forthcoming discussions the details of the theoretical results show, moreover, that the charged ends of the ligands are in the groove and do not exhibit direct interaction with the phosphates, a result confirmed by the X-ray analysis of reference (11). It may also be useful to add, for the same sake, that in the most refined computations, the difference in interaction energies of netropsin with the minor grooves of the AT and GC sequences, amounts to approximately 17 kcal/mole in favour of the former.

Because of the great importance of hydrogen bonds for the specific base pairings within the nucleic acids, the demonstration that the situation is more complex for the specificity of outside binding to these acids of external ligands is of obvious great interest. This interest was still strengthened by the lack of success of attempts which, within the classical concept of the decisive role of H-bonds, tried to modify the structure of the AT selective drugs, in particular netropsin, so as to make it GC specific.

Two proposals in this sense deserve a particular mention.

1) The first one due to Dickerson and collaborators (ref. 25) suggested that modifying netropsin by replacing its pyrrole(s) by imidazole(s), the ring nitrogen of which may form a hydrogen bond with the NH_2 group of guanine, could yield analogues capable of recognizing preferentially GC base pairs. Such analogues were named "lexitropsins". Fig. 2 indicates the chemical formulae and the denomination for three representative derivatives of this class.



Netropsin	X = CH, Y = CH
Lex A	X = N, Y = CH
Lex B	X = CH, Y = N
Lex AB	X = N, Y = N

Fig. 2. Lexitropsins

The theoretical exploration of this problem (ref. 26) did not confirm this qualitative suggestion. The computations show that although the complexation energy of the three lexitropsins with poly(dA).poly(dT) decreases progressively with the substitution of one and then two pyrroles by imidazoles and although the complexation energy of these ligands with poly(dG).poly(dC) increases progressively in the same circumstances, the

preference for the AT sequence is conserved for all the lexitropsins, although it diminishes with the number of imidazoles incorporated in place of pyrroles. Furthermore, details of the computations show that the dominance of the AT sequence over the GC sequence in all these interactions is due essentially to the greater value of the electrostatic component of the interaction energy with the former, reemphasizing thus the significance in this "specificity" of the stronger concentration of the molecular electrostatic potential in its minor groove. Obviously the new hydrogen bonds of lexitropsins are incapable of compensating completely for the initial electrostatic advantage (17 kcal/mole) of the minor groove of AT over that of GC sequences.

An experimental exploration of the binding and specificities of such lexitropsins (refs. 27,28) fully agreed with the above theoretical analysis and substantiates thus the important role in groove binding interactions of the electrostatic molecular potential.

2) The second proposal due to Goodsell and Dickerson (ref. 29) concerns another group of related ligands called isolexins which are pyrrole-amine and pyrrole-ketone analogs of netropsin, in which the netropsin backbone is shortened by eliminating either the C=O or the N-H group of the amide units in order to make them isohelical with DNA. It was again suggested that by suitably placing pyrrole or furan rings in these ligands, systems could be obtained capable to decipher the desirable DNA fragments, by appropriate hydrogen bonds between the H-bond donor and acceptor on the two interacting entities, following the scheme illustrated in Fig. 3.

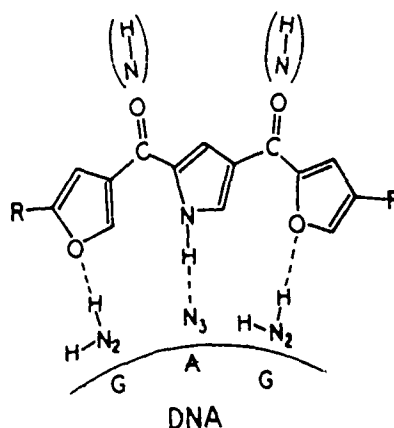


Fig. 3. Isolexins designed to bind to the GAG sequence.

It must be underlined that in fact the isolexins not only produce a shortening of the ligand to dimensions isohelical with DNA but imply also a substantially modified scheme of interactions with this biopolymer: this scheme involves now only the heteroaromatic rings (and possibly the end groups) but no more the groups bridging the rings (as was the case in

netropsin) which, whether C=O or N-H, are directed towards the exterior of the complex.

The detailed theoretical exploration of this system has led to a series of new, particularly striking results (refs. 30,31) which show that the potential specificity of isolexin type drugs for AT or GC minor groove binding depends to a large extent if not decisively, on a much greater number of factors than the formal possibilities of hydrogen bond formation as drawn in Fig. 3, some of which are most clearly understandable by the quantum chemical approach.

Thus :

1) The dominant role of the high electrostatic molecular potential in the minor groove of AT sequence is again demonstrated by the fact that when both end groups R of the isolexin of Fig. 3 are cationic (e.g. $R = -CH_2-CH_2-C(NH_3)^+$), this "furan-pyrrole-furan" isolexin although planned to bind to a GAG sequence, prefers to bind in fact to the regular AAA triplet.

2) When the R groups are neutral (e.g. $R = CH_3$) a particularly striking phenomenon happens which provides a strikingly useful illustration of the significance of quantum-chemical notions in the correct description of this type of interaction. Thus in the proposal of Goodsell and Dickerson (ref. 29) the only important factor is the shortening of the amide linkage and no distinction is made whether this occurs by leaving the C=O or N-H group. The nature of these groups (to which we shall refer as linkers), oriented towards the outside and not taking thus any direct part in the binding with DNA, was not considered as having a possible significance for the variation of specificity.

That this could be but a priori need not be the situation is evident if one considers the differences in the electronic properties of the C=O and N-H groups, in particular when they are engaged in conjugation (resonance) with π electronic systems, as it is the case in isolexins. Quantum chemistry (see e.g. ref. 32) teaches us that while $>C=O$ subtracts π electrons from such systems, $>N-H$ provides them with an excess of such electrons. It also indicates, that considering both the π and σ electrons, the C=O bond is associated with a dipole moment of about 2.4 D, with the overall direction $C^{\delta+}O^{\delta-}$, while the N-H bond is associated with a dipole moment of only about 1.3 D, with the overall direction $N^{\delta-}H^{\delta+}$. This situation must lead to significantly different dipole moments, both in magnitude and in direction, in isolexins utilizing the C=O or the N-H linkers. The natural question then arises : could these different electronic properties of the linking C=O or N-H groups of isolexins have any significant effect on the binding efficiencies and specificity in the interaction of these compounds with DNA ?

The answer to this question is provided by theoretical computations (ref. 30) which clearly show the significance of this situation : thus while the neutral "furan-pyrrole-furan" isolexin with two C=O linkers continues to prefer binding to the pure AT sequence, the ligand with two NH linkers shows a preference for the GAG triplet.

This striking reversal of the classical situation (AT specificity) is the more impressive as, obviously, it is related to the nature of the linking groups, C=O or N-H, which, as remarked above, are oriented externally with

respect to the oligonucleotide and do not take part directly in the binding process.

Inspection of the results shows that this different behaviour of the two lexitropsins springs essentially from the different evolution of the relatively small but decisive electrostatic component of their DNA-ligand interaction energy. A more elaborate interpretation can be provided in terms of the interaction of the adversely oriented C=O and N-H dipoles with the field of the oligonucleotide acceptors (ref. 30).

3) The significance of the linking groups in this process has still been enhanced by the demonstration (ref. 31) that the GC specificity of isolexins may be significantly increased by using as linkers the neutral -C=C- groups (yielding substances which we called vinylexins) to the point that the neutral vinylexins of Fig. 4 composed of three proton acceptor pentagonal heteroaromatic rings have been proposed as appropriate for binding to a homogeneous GGG sequence.

4) The efficiency and specificity of binding depends also on the nature of these heteroaromatic rings, the imidazole ring (Fig. 4b) being from that point of view superior to furan (Fig. 4a).

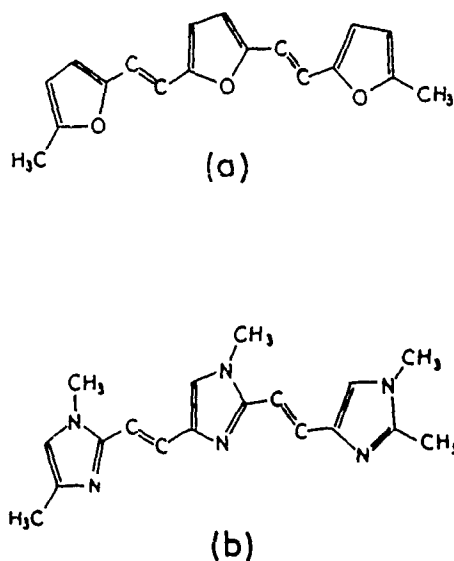


Fig. 4. Vinylexins designed to bind to the GGG sequence.

5) Finally, a complementary demonstration of the role of the end groups was given (ref. 31) by showing the enhancing effect upon both the specificity and the binding energy of the replacement of one of the terminal neutral groups of vinylexins by a cationic substituent.

The studies brought thus into evidence the insufficiency of considerations based solely on the notion of geometrical fitting and hydrogen bonding capabilities for a correct estimation of GC versus AT specificity of

groove binding ligands. Because of the existence in B-DNA of an intrinsic bias in favour of the minor groove of AT sequences with respect to GC sequences, due to the concentration of a deeper electrostatic potential in the former, correct evaluations of binding specificities can only be made by taking into consideration the overall electronic properties of the interacting species and explicitly calculating the energetics of DNA complex formation including all the relevant contributions. In these considerations and calculations the quantum mechanical concepts and methods play an essential role.

To terminate it may be interesting to add very recent experimental supports for some of the main proposals developed here. Thus the favouring effect for the GC specificity of the reduction of the positive charge of groove binding ligands, from two to one, was evidenced recently by a study of appropriate monocationic lexitropsins (ref. 33). Similarly the favouring effect in this respect of vinyl linkers was illustrated in a study by Baguley (ref. 34) of a series of bisquaternary ammonium heterocycles of the type of Fig. 5, in which this author has shown that the replacement of an amide linkage at position B or of an imine linkage at position A by a vinyl decreases strongly AT selectivity of this type of drugs. There was, of course, no chance that these dicationic compounds could show GC specificity but the decrease of AT specificity due to the vinyl linkers substantiates our hope that in an appropriate skeleton, such as that of the monocationic vinyllexins, they could produce GC specific ligands. Inversely one may say that Baguley's results show the particularly strong AT favouring effect of the amide linkage, which agrees well with our demonstration (ref. 30) of the significance in this respect of the interaction of the dipole moment of this linkage with the electrostatic field of DNA.

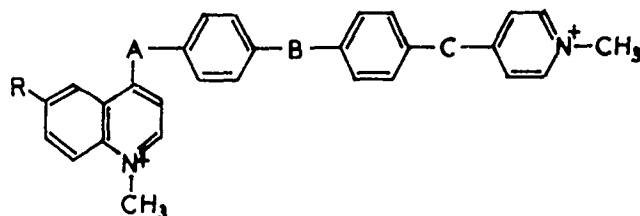


Fig. 5. Bisquaternary ammonium heterocycles of Baguley.

Finally, I would like to take advantage of this occasion to express my opinion on a surprising statement found recently in "Guidelines for Publications in Molecular Modeling related to Medicinal Chemistry" (ref. 35) Section g of the introduction states : "As previously recommended by the IUPAC with respect to QSAR studies, authors should REFRAIN FROM PUBLISHING PREDICTED ACTIVITIES OF SPECIFIC UNKNOWN STRUCTURES, since that may compromise the patentability of such structures if they are active, so there

is reduced incentive to synthesize these materials". To the annoyance of my students and collaborators upon reading this prescription I have first replied that on the contrary they should feel happy with it because if this situation is true they are free now to publish any crazy prediction they wish, as nobody is going to try to verify it anyway. More seriously, it seems astounding that anybody who believes that he has, after painstaking efforts, found a structure which predictably may be of importance for chemotherapeutic activity (say in cancer) should be advised to keep it in his drawers, in order not to prevent a hypothetical experimentalist to stumble upon it, possibly accidentally, in some unknown future. The cynicism of the "Guidelines" is in this respect shocking and the attitude which they represent immoral. Unless we consider seriously an advice by Crick (ref. 1 p.110) who writes : "The best a theorist can hope to do is to point an experimentalist in the right direction, and this is often best done by suggesting what directions to avoid". Taking as an example our theoretical search for GC minor groove specific isolexins described above, this would mean that we should have published only our negative results in this respect obtained for C=O linkers and dicationic end groups stressing that we only publish the negative ones. Leaving thus to a prospective experimentalist the chance to deduce from such a presentation the possible positive results which could be obtained with N-H linkers and neutral or monocationic end groups. Such a procedure could preserve our right to a courtously hidden intellectual priority and at the same time the right of the prospective experimentalist to the possible benefits of a successful patent. Such a compromise requires, however, a confidence in the honesty and intellectual abilities of the latter, which few theorists are probably ready to risk.

ACKNOWLEDGMENT

The author expresses his deep thanks to the Association for International Cancer Research (Brunel and St.Andrews Universities, United Kingdom) for its support of this work.

REFERENCES

- 1 F. Crick, *What Mad Pursuit*, Basic Books Inc., New York, 1988.
- 2 B. Pullman and A. Pullman, Molecular orbital calculation on the conformation of amino acid residues of proteins, *Advances in Protein Chemistry*, 28 (1974) 347 .
- 3 J.D. Watson, *The Double Helix*, Atheneum, New York, 1968.
- 4 R.B. Corey, The crystal structure of diketopiperazine, *J. Am. Chem. Soc.*, 60 (1938) 1598.
- 5 A. Pullman and B. Pullman, Electronic aspects of purine tautomerism, *Advances in Heterocyclic Chemistry*, 13 (1971) 77.
- 6 J.S. Kwiatkowski and B. Pullman, Tautomerism and electronic structure of biological pyrimidines, *Advances in Heterocyclic Chemistry*, 18 (1975) 199.

- 7 B. Pullman, Molecular mechanisms of specificity in DNA-antitumor drug interactions, *Advances in Drug Research*, 18 (1989) 1.
- 8 Ch. Zimmer and U. Wahnert, Nonintercalating DNA-binding ligands : specificity of the interaction and their use as a tool in biophysical, biochemical and biological investigations of the genetic material, *Progress Biophys. Molec. Biol.*, 47 (1986) 31.
- 9 G.V. Gursky, V.G. Tumanyan, A.S. Zasedatelev, A.L. Zhuze, S.L. Grokhovsky and B.P. Gottikh, A code governing specific binding of regulatory proteins to double helical DNA and RNA, in: *Nucleic Acid-Protein Recognition* (Ed), H.J. Vogel, Academic Press, 1977, pp. 189.
- 10 A.J. Krylov, S.L. Grokhovsky, A.S. Zasedatelev, A.L. Shuze, G.V. Gursky and B.P. Gottikh, Quantitative estimation of the contribution of pyrrolocarboxamide groups of the antibiotic distamycin A into specificity of its binding to DNA AT pairs, *Nucl. Acids Res.*, 6 (1979) 289.
- 11 M.L. Kopka, Ch. Yoon, D. Goodsell, P. Pjura and R. Dickerson, The binding of an antitumor drug to DNA : netropsin and CGCGAATT BrCGCG, *J. Mol. Biol.*, 183 (1985) 553.
- 12 M. Coll, J. Aymani, G.A. van der Marel, J.H. van Boom, A. Rich and A.H.J. Wang, A bifurcated hydrogen bonding conformation in the d(AT) base pairs of the DNA dodecamer d(CGCGAATTGCG) and its complex with distamycin, *Biochemistry*, 28 (1989) 310.
- 13 D.J. Patel and L. Shapiro, Sequence-dependent recognition of DNA duplexes : netropsin complexation to the TATA site of the d(GGTATACC) duplex in aqueous solution, *Biopolymers*, 25 (1986) 707.
- 14 R.E. Klevit, D.E. Wemmer and B.R. Reid, ¹H NMR studies on the interaction between distamycin A and a symmetrical DNA dodecamer, *Biochemistry* 25 (1986) 3296.
- 15 A.W. Braithwaite and B.C. Baguley, Existence of an extended series of antitumor compounds which bind to deoxyribonucleic acid by nonintercalative means, *Biochemistry*, 19 (1980) 1101.
- 16 B.C. Baguley, Nonintercalative DNA-binding antitumor compounds, *Molecular and Cellular Biochemistry*, 43 (1982) 167 .
- 17 B. Pullman and A. Pullman, Structural factors involved in the binding of netropsin and distamycin A to nucleic acids, *Studia Biophysica*, 86 (1981) 95.
- 18 A. Pullman and B. Pullman, Molecular electrostatic potential of nucleic acids, *Quart Rev. Biophys.*, 14 (1981) 289.
- 19 R. Lavery and B. Pullman, The dependance of the surface electrostatic potential of B-DNA on environmental factors, *J. Biomol. Struct. Dyn.* 2 (1985) 1021.
- 20 E. Scrocco and J. Tomasi, The electrostatic molecular potential as a useful tool for the interpretation of some molecular properties, *Top. Current Chem.*, 42 (1973) 91.
- 21 E. Scrocco and J. Tomasi, Electronic molecular structure, reactivity and intermolecular forces : an euristic interpretation by means of electrostatic molecular potentials, *Adv. Quantum Chem.*, 11 (1978) 115.
- 22 K. Zakrzewska, R. Lavery and B. Pullman, Theoretical studies of the selective binding to DNA of two non-intercalating ligands : netropsin and SN 18071, *Nucl. Acids Res.* 11, 8825 (1983).
- 23 R. Lavery, K. Zakrzewska and B. Pullman, Binding of non-intercalating antibiotics to B-DNA : a theoretical study taking into account nucleic acid flexibility, *J. Biomol. Struct. Dyn.* 3 (1986) 1155.
- 24 K. Zakrzewska, R. Lavery and B. Pullman, The solvation contribution to the binding energy of DNA with non-intercalating antibiotics, *Nucl. Acids Res.* 12 (1984) 6559.

- 25 M.L. Kopka, Ch. Yoon, D. Goodsell, P. Pjura and R.E. Dickerson, The molecular origin of DNA-drug specificity in netropsin and distamycin, *Proc. Natl. Acad. Sci. USA* 82 (1985) 1376.
- 26 K. Zakrzewska, R. Lavery and B. Pullman, A theoretical study of the sequence specificity in binding of lexitropsins to B-DNA, *J. Biomol. Struct. Dyn.* 4 (1987) 833.
- 27 J.W. Lown, K. Krowicki, U.G. Bhat, A. Skorobogaty, B. Ward and J.C. Dabrowiak, Molecular recognition between oligopeptides and nucleic acids : novel imidazole containing oligopeptides related to netropsin that exhibit altered DNA sequence specificity, *Biochemistry*, 25 (1986) 7408.
- 28 Ch. Zimmer, G. Luck, G. Burckhardt, K. Krowicki and J.W. Lown, The molecular mechanism of interaction of non-intercalative groove binding antitumor drugs with DNA, in *Molecular Mechanism in Carcinogenic and Antitumor Activity* (Ed). B. Pullman and C. Chagas, The Vatican Press, 1987, pp.339. Distributed by Adenine Press, New York.
- 29 D. Goodsell and R.E. Dickerson, Isohelical analysis of DNA groove-binding drugs, *J. Med. Chem.* 29 (1986) 727.
- 30 K. Zakrzewska and B. Pullman, Theoretical study of the sequence selectivity of isolexins, isohelical DNA groove binding ligands. Proposal for the GC minor groove specific compounds, *J. Biomol. Struct. Dyn.* 5 (1988) 1043 .
- 31 K. Zakrzewska, M. Randrianarivelo and B. Pullman, Drug recognition of DNA : proposals for GC minor groove specific ligands : vinylexins, *J. Biomol. Struct. Dyn.* 6 (1988) 331.
- 32 B. Pullman and A. Pullman, "Les Theories Electroniques de la Chimie Organique", Masson (Ed.), Paris, 1952.
- 33 K. Kissinger, K. Krowicki, J.C. Dabrowiak and J.W. Lown, Molecular recognition between oligopeptides and nucleic acids. Monocationic imidazole lexitropsins that display enhanced GC sequence dependent DNA binding, *Biochemistry* 26 (1987) 5590.
- 34 B. Baguley, private communication.
- 35 P. Gund, D.C. Barry, J.M. Blaney and N.C. Cohen, Guidelines for publications in molecular modeling related to medicinal chemistry, *J. Med. Chem.*, 31 (1988) 2230.

DISCUSSION

MAROUN - For hexitropsin AB, and for the neutral isolexins, for example, your calculations show a complexation energy difference between the AT and the GC sequence of ~3.5 Kcal/mole. In other instances, the differences amount to 29 Kcal/mole. Are the quantities on the former cases significant enough to draw conclusions concerning the sequence specificities of this family of compounds ?

B. PULLMAN - I believe that I have given the answer to this question in my very talk. Because we thought that a difference of specificity of 3.5 Kcal/mole in our prototype was not significant enough (although significant nevertheless) we worked on improving it by looking for structural modifications of the Ligands which would just do that and we succeeded. It is certain that small indices of specificity are less reliable than great ones. By great ones I mean equal or above 6-7 Kcal/mole.

GENERAL THEORY OF INTERMOLECULAR FORCES

A.D. BUCKINGHAM

University Chemical Laboratory, Lensfield Road, Cambridge, CB2 1EW, U.K.

SUMMARY

There is much current interest both in inter- and intra-molecular potential energy surfaces. The structure and properties of van der Waals molecules and clusters provide an important source of information about molecular interactions. This information can be used to generate intermolecular potentials that can be useful in descriptions of larger systems, such as condensed phases, solutions of surfactants, and biomacromolecules. It is convenient to divide intermolecular potentials into long-range and short-range components. The former are related via perturbation theory to the charge distribution and polarizabilities of the free molecules, and the resulting long-range potentials vary as an inverse power of the separation between the molecules. The short-range interactions result from the overlap of the electron clouds of the interacting molecules and diminish exponentially with the separation. By dividing the molecules into atoms or small groups of atoms, it is possible to obtain convenient and convergent representations of the potential. The approach can be used to provide a theoretical basis for the popular site-site potential models that are now used extensively in computer simulations. Attention will be paid to the additivity or non-additivity of potentials, to the rôle of the solvent in solutions, and to changes in the electronic properties, such as the dipole moment, that result from intermolecular forces.

INTRODUCTION

Intermolecular forces have an important rôle in many branches of science, including physics, chemistry, molecular biology, materials science, and crystallography. Attractive forces are responsible for the existence of liquids and solids and for the structure of biomacromolecules. The repulsive forces that operate through the overlap of electron clouds when atoms are close together determine nearest-neighbour distances and the densities of liquids and solids as well as their compressibility. Much quantitative information about intermolecular potentials is available and in most cases this has been obtained through the study of small molecules, and in particular from pairwise interactions, as in a van der Waals dimer. This accurate information can be applied to the study of larger molecules, particularly through dividing the large molecule into smaller parts involving groups of atoms that occur in simpler systems.

Even for small molecules, difficulties exist. For example, the interaction of two water molecules in the water dimer $(\text{H}_2\text{O})_2$ involves a potential surface depending on twelve coordinates - these are the distance R between the centres of mass, the angles θ_a and θ_b between the dipole axes μ_a and μ_b of each H_2O and the intermolecular vector R , the azimuthal angle ϕ between the planes containing μ_a and R and μ_b and R , the orientations χ_a and χ_b of each H_2O about its dipole axis, together with the three internal vibrational degrees of freedom in each H_2O

molecule, involving the OH bond lengths and the HOH angle. Thus even in this case simplifications are essential. In the general case of interacting molecules containing N_a and N_b nuclei, there are $3(N_a + N_b) - 6$ independent variables (except when $N_a = N_b = 1$ when there is just one variable which is the internuclear distance R).

Of these $3(N_a + N_b) - 6$ degrees of freedom ($3N_a - 6$) and ($3N_b - 6$) are associated with the internal coordinates of the molecules a and b and the remaining 6 describe the relative positions and orientations of the two molecules - the so-called intermolecular degrees of freedom.

Fortunately, there is an important simplification and this is that the change in the molecular structure and properties as a result of the interaction is generally small. We can identify the molecule a and the molecule b in the interacting system, and we can use perturbation theory and our knowledge of free a and free b to describe the interaction of a and b. This perturbation approach is valid provided the effects of intermolecular exchange are small, and that is the case in the 'long-range' region. Overlap of the electron clouds causes a loss of 'identity' of the molecules - it is associated with the pair ab and limits our ability to relate the interaction to the properties of free a and b.

The Hamiltonian of an interacting pair a,b is

$$H = H_a + H_b + H_{int} \quad (1)$$

where H_a and H_b are the Hamiltonians for free a and free b and

$$H_{int} = (4\pi\epsilon_0)^{-1} \sum_{i,j} e_i^{(a)} e_j^{(b)} / R_{ij} \quad (2)$$

is the Coulombic interaction of the charges of molecules of a and b; R_{ij} is the distance between the i th charge $e_i^{(a)}$ in molecule a and the j th charge $e_j^{(b)}$ in b.

In the general theory of long-range intermolecular forces (refs. 1-2), H_{int} is treated as a perturbation to $H_a + H_b$, so the unperturbed basis is non-interacting a and b. The unperturbed wavefunction in the non-degenerate case is a product $\psi_{m_a} \psi_{m_b}$ where ψ_{m_a} and ψ_{m_b} are the vibronic states of free a and free b. Generally ψ_{m_a} and ψ_{m_b} will be the ground states but sometimes we are interested in excited electronic or vibrational states.

The perturbed wavefunction is

$$\psi_{m_a m_b} = \psi_{m_a} \psi_{m_b} + \sum_{p_a, p_b} \frac{\langle \psi_{p_a} \psi_{p_b} | H_{int} | \psi_{m_a} \psi_{m_b} \rangle}{E_{m_a} + E_{m_b} - E_{p_a} - E_{p_b}} \psi_{p_a} \psi_{p_b} + \dots \quad (3)$$

where $\sum_{p_a p_b}'$ denotes a sum over the complete set of unperturbed states $\psi_{p_a} \psi_{p_b}$ with the exception of the initial state $\psi_{m_a} \psi_{m_b}$. The first-order perturbed wavefunction therefore consists of the addition to the unperturbed function of a small amount of all those unperturbed states of free a and b that are mixed with $\psi_{m_a} \psi_{m_b}$ by the perturbation H_{int} . The extent of the admixture is given by eqn (3).

The energy of the pair is

$$E_{m_a m_b} = \langle \psi_{m_a m_b} | H | \psi_{m_a m_b} \rangle / \langle \psi_{m_a m_b} | \psi_{m_a m_b} \rangle \quad (4)$$

$$= E_{m_a} + E_{m_b} + \langle \psi_{m_a} \psi_{m_b} | H_{int} | \psi_{m_a} \psi_{m_b} \rangle - \sum_{p_a, p_b}' \frac{|\langle \psi_{p_a} \psi_{p_b} | H_{int} | \psi_{m_a} \psi_{m_b} \rangle|^2}{E_{p_a} - E_{m_a} + E_{p_b} - E_{m_b}} + \dots \quad (5)$$

The first-order energy in (5) is the unperturbed expectation value of H_{int} and is the electrostatic energy $u_{electrostatic}$.

The second-order energy may be separated into two distinct contributions. The first, $u_{induction}$, consists of all those terms in which either $p_a = m_a$ with $p_b \neq m_b$, or $p_b = m_b$ with $p_a \neq m_a$. The other second-order contribution is $u_{dispersion}$ and it is comprised of the remaining terms in which $p_a \neq m_a$ and $p_b \neq m_b$:

$$u_{induction} = u_{induction}^{(a)} + u_{induction}^{(b)} \quad (6)$$

where

$$u_{induction}^{(a)} = - \sum_{p_a \neq m_a} \frac{|\langle \psi_{p_a} \psi_{m_b} | H_{int} | \psi_{m_a} \psi_{m_b} \rangle|^2}{E_{p_a} - E_{m_a}} \quad (7)$$

$$u_{induction}^{(b)} = - \sum_{p_b \neq m_b} \frac{|\langle \psi_{m_a} \psi_{p_b} | H_{int} | \psi_{m_a} \psi_{m_b} \rangle|^2}{E_{p_b} - E_{m_b}} \quad (8)$$

$$u_{dispersion} = - \sum_{\substack{p_a \neq m_a \\ p_b \neq m_b}} \frac{|\langle \psi_{p_a} \psi_{p_b} | H_{int} | \psi_{m_a} \psi_{m_b} \rangle|^2}{E_{p_a} - E_{m_a} + E_{p_b} - E_{m_b}} \quad (9)$$

If ψ_{m_a} and ψ_{m_b} are ground electronic states, then $u_{\text{dispersion}}$ is necessarily negative, leading to an attractive force between a and b; $u_{\text{induction}}^{(a)}$ is also negative if molecule b in its unperturbed state produces a non-vanishing electric field at a, and similarly for $u_{\text{induction}}^{(b)}$.

If $\psi_{m_a m_b}$ is degenerate, then it is necessary to select the unperturbed zero-order wavefunction that diagonalizes H . The degeneracy may be lifted by H_{int} , in which case there is a resonance energy $u_{\text{resonance}}$ which may be either attractive or repulsive. The first-order perturbed energy is now the sum of $u_{\text{electrostatic}} + u_{\text{resonance}}$. For example, consider an atom a in an excited state p_a interacting with an identical atom b in its ground state m_b . The appropriate zeroth-order states are (ref. 2)

$$\psi_g = (\psi_{p_a} \psi_{m_b} + \psi_{m_a} \psi_{p_b}) / \sqrt{2} \quad (10)$$

and

$$\psi_u = (\psi_{p_a} \psi_{m_b} - \psi_{m_a} \psi_{p_b}) / \sqrt{2} \quad (11)$$

and there is an exchange of excitation energy due to H_{int} which splits the energies of these g and u states by $2 \langle \psi_{p_a} \psi_{m_b} | H_{\text{int}} | \psi_{m_a} \psi_{p_b} \rangle$. If m is an S state and p a P state, then the longest-range contribution to $u_{\text{resonance}}$ will vary as the inverse cube of the distance between the atoms, due to a dipolar exchange of a photon. If p were an excited D state, then the longest-range resonance interaction would vary as R^{-5} due to quadrupolar exchange.

An interesting example of $u_{\text{resonance}}$ occurs in vibrationally excited van der Waals molecules A_2 . Consider the hydrogen fluoride dimer $(\text{HF})_2$ in which one of the HF monomers is vibrationally excited. The interaction of the monomers means that the vibrational excitation can be resonantly transferred between a and b, leading to a splitting of the potential surface. $(\text{HF})_2$ has a well-known inversion doubling of its energy levels due to the tunnelling motion that interchanges the proton donor and acceptor in the hydrogen-bonded dimer (ref. 3). The

tunnelling frequency is about 20 GHz and it is reduced to about one-third this value in the excited vibrational states $\nu_1 = 1$ at 3930.9 cm^{-1} and in $\nu_2 = 1$ at 3868.3 cm^{-1} (ref. 4). The barrier height opposing the tunnelling motion is apparently increased from 300 cm^{-1} to about 400 cm^{-1} in $\nu_1 = 1$ and 391 cm^{-1} in $\nu_2 = 1$ (ref. 4). This change in the effective potential energy surface with vibrational excitation can be attributed to $u_{\text{resonance}}$ (ref. 5). It will affect $(\text{HF})_2$ but not (HFDF) , since there can be no resonance transfer in the latter case, so this mass-dependent potential-energy surface goes beyond the usual clamped-nuclei Born-Oppenheimer picture.

In the ground state the deuterium-bonded species $\text{HF}\cdots\text{DF}$ is favoured over the hydrogen-bonded $\text{DF}\cdots\text{HF}$ (ref. 3,6) and it has been calculated that this energy difference, resulting from zero-point vibrations and in particular the bending of the D and H bonds, is about 163 cm^{-1} (ref. 7). However, in the excited vibrational levels the species HFDF will not experience a resonance energy as in $(\text{HF})_2$ and $(\text{DF})_2$.

It is convenient to classify long-range intermolecular interactions as in table 1, which

TABLE 1.

A classification of molecular interaction energies

Range	Type	Attractive(-) or repulsive (+)	Additive or nonadditive
long	electrostatic	-/+	additive
	induction	-	nonadditive
	dispersion	-	approximately additive
	resonance	-/+	nonadditive
	magnetic	-/+	(weak)
short	overlap (Coulomb and exchange)	-/+	nonadditive

indicates whether a particular interaction is additive or non-additive depending on whether the energy of a cluster of three molecules can or cannot be represented as a sum of the three pair-interactions (ref. 2).

The non-additive induction energy can be important in solvent effects on intermolecular forces between ions or between highly polar species. The induction energy is determined by the square of the electric field strength at the polarizable site and it will therefore cause an attractive force between ions of the same sign and a repulsion between ions of opposite sign (ref. 8).

Limitations of the long-range description of intermolecular forces

The great advantage of the long-range description of molecular interactions is that it is a general theory that expresses the interactions in terms of the properties of the isolated molecules. The long-range energy or property, such as the induced dipole moment, is obtained in any particular case by inserting the appropriate monomer properties into the general equations, giving the dependence of the energy or property on the intermolecular degrees of freedom. The monomer charge distributions determine the electrostatic energy, and their static and dynamic polarizabilities determine the induction and dispersion energies, respectively (ref.2). Some of these properties are accurately known from experiment, and others are amenable to *ab initio* computation.

But what are the limitations to this description? At what separations R do the overlap effects cause significant departures from the long-range formulation? In an attempt to answer this important question, we carried out a series of elaborate *ab initio* computations on the system Ne HF (ref. 10). The calculations were performed within the self-consistent field (SCF) framework, using large basis sets; the results are, therefore, close to the Hartree-Fock limit, but they exclude any effects of electron correlation (so dispersion forces are absent). An earlier calculation on Ne HF by Losonczy *et al.* (ref. 11) had led to the conclusion that covalency (a short-range overlap effect) made a dominant contribution to the energy at the potential minimum

and that long-range theory was applicable only when the separation R of the Ne and HF molecules is so large that the interaction is negligible.

However, the important role attributed to covalency in Ne HF (ref. 11) is actually an artefact due to basis set superposition error. By using the counterpoise technique of Boys and Bernardi (ref. 12), it has been shown (ref. 10) that, even down to the minimum-energy separation of Ne ... HF, the long-range theory accounts well for the interaction energy $U(R)$ and for the dipole moment $\mu(R)$.

The problem of computing the electrostatic and induction energies of van der Waals molecules has largely been overcome by using the method of distributed multipoles and distributed polarizabilities, introduced by Stone (refs. 13-14). This approach has yielded a reliable practical means of calculating the structures (ref. 15) and dipole moments (ref. 16) of van der Waals molecules. We are currently investigating whether this 'long-range' approach will give good dipole and polarizability derivatives, and hence infrared and Raman intensities.

The long-range forces are balanced by the repulsive short-range forces at equilibrium, and these are not amenable to evaluation by perturbation theory. Various simple models have been adopted (refs. 9,15) but there remains a need to obtain better representations of these interactions.

REFERENCES

- 1 H.C. Longuet-Higgins and L. Salem, Proc. Roy. Soc. A., 259 (1961) 433.
- 2 A.D. Buckingham, Adv. Chem. Phys., 12 (1967) 107.
- 3 T.R. Dyke, B.J. Howard and W. Klemperer, J. Chem. Phys., 56 (1972) 2442.
- 4 A.S. Pine, W.J. Lafferty and B.J. Howard, J. Chem. Phys., 81 (1984) 2939.
- 5 G.T. Fraser, J. Chem. Phys., 90 (1989) 2097.
- 6 H.S. Gutowsky, C. Chuang, J.D. Keen, T.D. Klots and T. Emilsson, J. Chem. Phys. 83 (1985) 2070.
- 7 A.D. Buckingham and Liu fan-chen, Int. Rev. Phys. Chem 1 (1981) 1953.
- 8 See the Faraday Discussion on 'Solvation', Faraday Discuss. Chem. Soc. 85 (1988) 78-83.
- 9 A.D. Buckingham, P.W. Fowler and J.M. Hutson, Chem. Revs., 88 (1988) 963.
- 10 P.W. Fowler and A.D. Buckingham, Mol. Phys., 50 (1983) 1349.
- 11 M. Losonczy, J.W. Moscovitz and F.H. Stillinger, J. Chem. Phys., 61 (1974) 2438.
- 12 S.F. Boys and F. Bernardi, Mol. Phys., 19 (1970) 553.
- 13 A.J. Stone, Chem. Phys. Letters, 83 (1981) 233.
- 14 A.J. Stone, Mol. Phys., 56 (1985) 1065.
- 15 A.D. Buckingham and P.W. Fowler, Canadian J. Chem., 63 (1985) 2018.
- 16 A.D. Buckingham, P.W. Fowler and A.J. Stone, Int. Rev. Phys. Chem., 5 (1986) 107.

DISCUSSION

SOUMPASIS - Prof. Buckingham's example of the interaction of two ions and a solvent molecule applies in isolated systems (few particles). In large systems the dielectrically screened Coulomb interaction (continuum model) is capable of describing many things due to statistical averaging of the detailed interactions.

BUCKINGHAM - I agree that my example of a polarizable atom at the midpoint of (a) an anion and a cation and (b) two cations represents a single configuration and therefore excludes the influence of entropy. It should not be taken to represent ionic interactions in an aqueous solution. Nevertheless, it illustrates the deficiency of the dielectric continuum model, which would treat (a) and (b) similarly and which always reduces the magnitude of the interaction between the ions. In case (a) the interaction between the ions is actually enhanced by the presence of a water molecule or polarizable atom between the charges.

RIVAIL - Could you comment a bit further on the transferability of atomic or bond quantities which govern intermolecular energies from one molecular species to another one and, occasionally, from intermolecular to intramolecular energies (in the case of macromolecules).

BUCKINGHAM - I believe that accurate experiments and computations on small systems, particular Van der Waals molecules, provide the firm basis upon which we can build our approximate descriptions of the interactions in macromolecules and between large molecules. The parameters describing the charge distribution and polarizabilities of atoms and/or groups of atoms are likely to be similar in appropriate cases, so we should be able to gain at least a qualitative understanding of the potential energy surface in a particular case from a knowledge of the corresponding parameters in small molecules. For example, as a first guess for the distributed multipoles in a carbonyl-containing molecule, one could use the values computed for formaldehyde. You have raised a large and very important question that needs careful and extensive investigation.

DURUP - In the case of two cations with two polarizable molecules, where second-order effects are important and where a set of other solvent molecules has to be considered, how large should be the whole system considered for the theoretical treatment to converge ?

BUCKINGHAM - There are two aspects to your question : the first is how large a sample needs to be to reproduce the continuum results ? For two actions the convergence in the case of a spherical sample will go as R^{-1} and for an uncharged (cation + anion) it will be faster and go as R^{-3} , where R is the radius of the sphere. The second part asks how high in powers of the polarizability α one has to go to achieve convergence. It depends on $\alpha(4\pi\epsilon_0 R^3)^{-1}$ where R is the distance of the atom from the charges. For a linear system with charges q at $\pm R_0$ and polarizable atoms at $\pm R$ the induction energy can be obtained exactly as

$$-\alpha \left[\frac{2q(R^2 + R_0^2)}{(4\pi\epsilon_0)(R^2 - R_0^2)^2} \right]^2 \left[1 + \frac{\alpha}{4R^3(4\pi\epsilon_0)} \right]^{-1}$$

Hence the power series in the polarizability converges fast when $\alpha/(4\pi\epsilon_0)4R^3 \ll 1$, but will not converge when $\alpha/(4\pi\epsilon_0)4R^3 \rightarrow 1$, even though an exact analytical expression exists. In this case, the higher-order interactions diminish the induction energy since the fields of the induced dipoles oppose the field of the charges (q).

DURUP - For the HF dimer, does the tunneling frequency reincrease if both HF molecules have one vibrational quantum ?

BUCKINGHAM - I do not think that the experiment has been done. However, as you imply, the resonance energy will not be present if both HF monomers are excited into $v=1$. This resonance energy is large when one HF has $v=1$ and the other $v=0$. It will be small when one has $v=2$ and the other $v=0$, and is zero when both have 0,1,2,... quanta.

SQUIMPASIS - The effects of various Hamiltonians in describing many body properties of classical charged fluids (e.i. electrolytes) have been studied in some detail. The asymptotically correct (long range) effective potential of two ions in the polar solvent (e.i. water) is a dielectrically screened Coulomb form as assumed in the restricted primitive model (ions : charged hard spheres, water : dielectric continuum). More "realistic" solvent modelling (water : hard dipolar spheres) yields a different effective interaction. But still more realistic modelling (water : hard dipolar + quadrupolar spheres) yields effective potentials identical to the dielectric screened Coulomb over most distances. The statistical mechanics of dense charged systems are very often dominated by the long range part of the effective potential and therefore refinements at very short distances (induction, dispersion, etc.) often do not give rise to dramatic improvements in the description of the condensed phase.

BUCKINGHAM - The point of my simple example of a polarizable particle near two charges was to emphasize the deficiency in the dielectric screening description at short range - it may give the wrong sign. At very long range, the dielectric screening correctly describes the interaction of fixed, or slowly fluctuating, charges. I agree that this long-range behaviour is dominant in determining some properties. However, the structures of flexible molecules containing charges and dipoles are dependent upon near-neighbour interactions where the dielectric screening approximation fails.

THEORETICAL MODELS FOR INTERMOLECULAR POTENTIALS

A. J. STONE

University Chemical Laboratory,
Lensfield Road,
Cambridge CB2 1EW,
England

ABSTRACT

It is now possible to calculate the interaction energies of small molecular complexes by a variety of *ab initio* techniques, though there are several sources of error, notably Basis Set Superposition Error, which is especially troublesome when electron correlation is taken into account. Such calculations have to be carried out at a wide range of dimer geometries if a full description of the potential energy surface is needed, and this is extremely time-consuming. For complexes involving larger molecules it is out of the question.

An alternative approach is to isolate the components of the perturbation expansion of the interaction energy, namely the repulsion, electrostatic, induction, and dispersion terms, and to calculate each of them independently by the most appropriate technique. Thus the electrostatic interaction can be calculated accurately from distributed multipole descriptions of the individual molecules, while the induction and dispersion contributions may be derived from molecular polarizabilities. This approach has the advantage that the properties of the monomers have to be calculated only once, after which the interactions may be evaluated easily and efficiently at as many dimer geometries as required. The repulsion is not so amenable, but it can be fitted by suitable analytic functions much more satisfactorily than the complete potential. The result is a model of the intermolecular potential that is capable of describing properties to a high level of accuracy.

THE SUPERMOLECULE METHOD

A natural way to calculate the interaction energy U_{AB} between two molecules A and B is to compare the energy of the complex $A \cdots B$ with the energies of the separated molecules:

$$U_{AB} = W_{AB} - W_A - W_B. \quad (1)$$

This is the *supermolecule method*. It has a number of obvious attractions: it is very easy to understand conceptually, and it is easy to apply, using standard *ab initio* computational techniques. Unfortunately it has some serious limitations. The small interaction energy U_{AB} is calculated as a difference of much larger quantities, and this is a serious source of inaccuracy. We are all aware of the need to avoid calculating a small quantity as a difference of large ones for reasons of numerical accuracy, but the problem here is not a numerical one; it is that all of the energies involved are subject to error, because of the approximations made in the calculation, and small relative errors in any of the energies lead to very large percentage errors in the interaction energy. Moreover the errors in W_{AB} do not cancel

with the errors in W_A and W_B . Consequently the calculations must be carried out at a high level of accuracy if the results are to be of any value. However this means that a large basis set is needed, and since the computer time required for *ab initio* calculations increases roughly with the fourth power of the number of basis functions, these calculations are very expensive.

Furthermore, the calculation has to be repeated at a large number of relative configurations of the two molecules if the potential surface is to be explored adequately. Six coordinates are required to specify the position and orientation of molecule *B* relative to molecule *A*, if both are non-linear. If the calculation is repeated for only four values of each coordinate (a very inadequate number) a total of $4^6 = 4096$ points will be required. There are many examples in the literature of calculations of intermolecular potential energy surfaces which use a wholly inadequate strategy for choosing points on the surface. It is common, for example, even in calculations on pairs of diatomics, where there are only three angular coordinates, to perform calculations at only a few relative orientations: often the linear, rectangular, T-shaped and crossed configurations only. Even if calculations are done for a large number of intermolecular distances, it is impossible to characterise the potential energy surface adequately from such a limited number of orientations. Fortunately it is possible to explore the surface much more efficiently[1], but a large number of points is still required.

It is often assumed that the supermolecule method provides the ultimate method for calculating intermolecular interaction energies, and that it provides a benchmark for all other methods. This is not entirely true. Supermolecule calculations are subject to a number of deficiencies, because they inevitably use inadequate basis sets and do not take full account of electron correlation.

A complete treatment of electron correlation, even within a moderate-sized basis set, is impracticable for all but the smallest calculations. It is therefore necessary to use a limited treatment. The simplest of these, the so-called 'singles-and-doubles' CI, or CISD, takes into account the contributions to the wavefunction of all singly and doubly-excited configurations derived from the SCF 'root' configuration. This is a variational technique, but unfortunately is not 'size-extensive': the energy obtained for a complex $A \cdots B$ at infinite separation is not the sum of the energies obtained for *A* and *B* individually[2]. The requirement of size extensivity is clearly important if sensible estimates of interaction energies are required. Although the CISD method does not satisfy this requirement, there are methods which do; they include Møller-Plesset perturbation theory and a variety of 'coupled-cluster' methods[2,3]. CEPA (the Coupled Electron Pair Approximation) is an approximation to the latter. However none of these is variational, and all are approximate.

The basis set problem is common to all *ab initio* calculations, but is particularly troublesome in the calculation of interaction energies, because even when the energies of *A*, *B* and $A \cdots B$ are calculated variationally, the difference (1) is not an upper bound to the interaction energy, and it is difficult to estimate the error in it. A well-known deficiency is 'Basis Set Superposition Error', or BSSE, which arises as follows.

We choose a basis for the calculation of molecule *A* that seems suitable for our purpose. Inevitably it is incomplete, so our wavefunction for *A* is not exact. The same is true of the calculation for the isolated molecule *B*. Now we carry out a calculation on the complex $A \cdots B$, and obtain a new energy that includes the interaction between the molecules. But in this calculation there are some new basis functions, belonging to molecule *B*, that were not present when we calculated the energy of the isolated *A* molecule, and they allow the wavefunction of molecule *A* to be improved variationally, so that its energy falls. This is quite separate from the true physical interaction; it is a spurious effect that occurs because the basis set that we initially chose for *A* was not good enough to describe the wavefunction exactly. This is Basis Set Superposition Error, and its effect is to make the interaction seem more attractive than it really is.

This is an extremely troublesome problem. The standard way to deal with it is the *counterpoise correction* proposed by Boys & Bernardi[4]. The reference energy for *A* is calculated in the presence of the basis functions (but not the electrons or nuclei) of molecule *B*, and similarly the reference energy for *B* is calculated in the presence of the *A* basis functions. In this way, the variational improvement that arises from the presence of the 'foreign' basis functions is included in the reference calculation as well as in the $A \cdots B$ calculation. (The basis functions of the other molecule are sometimes called 'ghost' orbitals when used like this.)

It has been pointed out that in the $A \cdots B$ complex some of the orbitals of *B* are not available to *A*, because they are occupied by the electrons of *B*, and it has been argued that in the counterpoise calculation these orbitals should be excluded. There has been a great deal of controversy about this point[5]. This has rather obscured the fact that even if the effects of BSSE could be corrected completely, the calculation would still be in error because of the inadequacy of the basis set and the correlation treatment. In a recent survey of calculations on the water dimer[6], it was concluded that the very best calculations currently available were in error by some 120 cm^{-1} . This is a substantial fraction of the interaction energy.

INTERMOLECULAR PERTURBATION THEORY

If supermolecule calculations are expensive and unreliable for calculations on small van der Waals complexes, it is clear that they are useless for calculating interactions involving large molecules such as proteins. However perturbation theory can provide a more satisfactory approach. It is often assumed that perturbation theory is a relatively crude approach to the calculation of intermolecular potentials, but there is no reason to believe that it is intrinsically any less accurate than the supermolecule approach. The attraction of perturbation theory is that it leads directly to an expression for the interaction energy itself. Moreover it is possible to analyse the expression into separate terms that can be correlated with distinct physical effects. We can then refine each of these terms, expressing each of them in terms of properties of the individual molecules that can be calculated much more accurately than properties of the complex. Unfortunately perturbation theory becomes very much more difficult at short range, so it is convenient to distinguish between this and the long-range case, where the overlap between the wavefunctions on different molecules can be neglected.

Long-range perturbation theory—first order

For two molecules, sufficiently far apart, the Hamiltonian for the system can be written as

$$\mathcal{H} = \mathcal{H}_A + \mathcal{H}_B + V, \quad (2)$$

where \mathcal{H}_A is the Hamiltonian for the isolated molecule *A*, \mathcal{H}_B the Hamiltonian for molecule *B*, and *V* is the interaction. *V* arises simply from the electrostatic interaction between the particles of *A* and those of *B*:

$$V = \sum_{i \in A, j \in B} \frac{e_i e_j}{4\pi\epsilon_0 r_{ij}}, \quad (3)$$

where e_i and e_j are the charges on particles *i* and *j*, and r_{ij} is the distance between them.

If the eigenfunctions of \mathcal{H}_A are ψ_m^A and those of \mathcal{H}_B are ψ_n^B , and if there is no overlap between the two sets of functions, then the unperturbed wavefunctions for $A \cdots B$ are $|mn\rangle = \psi_m^A \psi_n^B$. Ordinary Rayleigh-Schrödinger perturbation theory then gives the first-order energy of interaction as

$$W^{(1)} = \langle 00|V|00\rangle \quad (4)$$

$$= \int \frac{\rho^A(\mathbf{r}_1)\rho^B(\mathbf{r}_2)}{r_{12}} d^3r_1 d^3r_2, \quad (5)$$

which is just the classical electrostatic energy of interaction between the charge distributions of A and B . It is easily evaluated if the ground state wavefunctions are known, but the calculation requires a knowledge of all the intermolecular electron repulsion integrals and is fairly time-consuming. If we want to map out the potential energy surface, then it is necessary to repeat the calculation at a large number of relative configurations, just as in the supermolecule method, and this would require an inordinate amount of computer time if it were to be done accurately. The electrostatic interaction is quite sensitive to the effects of electron correlation, because the individual molecular charge distributions are modified quite significantly when electron correlation is taken into account.

It is often useful to replace the operator (3) by its *multipole expansion*:

$$V = \sum_{tu} \tilde{Q}_t^A T_{tu}^{AB} \tilde{Q}_u^B, \quad (6)$$

where \tilde{Q}_t^A is an operator for one of the multipole moments of molecule A (charge, dipole, quadrupole, etc.) and T_{tu}^{AB} is an interaction function. The index t is an abbreviation for the angular momentum labels lk , and the moment operator \tilde{Q}_{lk}^A is defined by

$$\tilde{Q}_{lk}^A = \sum_{i \in A} e_i r_i^l C_{lk}(\theta_i, \phi_i), \quad (7)$$

where r_i , θ_i , and ϕ_i are the spherical polar coordinates of particle i in a local coordinate system for molecule A , and C_{lk} is a modified spherical harmonic:

$$C_{lk}(\theta, \phi) = (4\pi/(2l+1))^{1/2} Y_{lk}(\theta, \phi).$$

The derivation of (6) is straightforward[7,8], and is not given here.

If now we use the multipole expansion of the perturbation in the calculation of the first-order energy, eq. (4) becomes simply

$$U_{es}^{AB} = \sum_{tu} Q_t^A T_{tu}^{AB} Q_u^B, \quad (8)$$

in which Q_t^A is the expectation value of one of the multipole moments of molecule A , referred to local axes, and the interaction function T_{tu}^{AB} takes into account the orientation dependence of the interaction as well as the distance between the molecules. The interaction is now described in terms of monomer properties (the multipole moments Q_t^A and Q_u^B) which need to be calculated only once, and so can be calculated at a much higher level of accuracy than we could contemplate for calculations on the complete complex. To obtain the interaction energy at any arbitrary configuration it is now necessary only to calculate the interaction functions T_{tu}^{AB} for that orientation, which is a trivial computation.

We note in passing that it is possible, and much more useful for practical calculations, to use an alternative approach in which the indices t and u refer to real multipole moments, rather than the complex ones defined by (7). The moments are now denoted Q_{lk_c} and Q_{lk_s} , defined, for $k > 0$, by

$$\begin{aligned} Q_{lk_c} &= \sqrt{\frac{1}{2}} [(-1)^k Q_{lk} + Q_{l,-k}], \\ iQ_{lk_s} &= \sqrt{\frac{1}{2}} [(-1)^k Q_{lk} - Q_{l,-k}]. \end{aligned} \quad (9)$$

No transformation is needed for Q_{l0} , which is always real. The notation reflects the fact that Q_{lk_c} transforms like $\cos k\phi$ and Q_{lk_s} like $\sin k\phi$. The factors of $\sqrt{1/2}$ ensure that a rotation of axes induces an orthogonal transformation of the moments. The first few of these moments coincide precisely with the Cartesian charge and dipole moment:

$$\begin{aligned} Q_{00} &= \sum_i e_i, \\ Q_{10} &= \sum_i e_i z_i = \mu_z, & Q_{11_c} &= \sum_i e_i x_i = \mu_x, & Q_{11_s} &= \sum_i e_i y_i = \mu_y, \end{aligned} \quad (10)$$

and later ones describe the quadrupole, octopole and so on. A complete list is given in ref. 7 for moments up to hexadecapole. The functions T_{tu}^{AB} for this formulation have been tabulated[7,9] for all multipole-multipole interactions up to terms in R^{-5} .

Although the expression (8) is accurate at sufficiently large separation, it converges only if the separation between the molecules is large compared with their size. This is not the case for any pair of molecules at the separations found in condensed phases or in weakly bound complexes. Fortunately this problem is easily overcome, provided that the molecular charge distributions do not overlap, by using a *Distributed Multipole Expansion*, which assigns charges, dipole moments, etc., to *regions* comprising single atoms or (in larger molecules) small groups of atoms. For small molecules it is helpful to treat the bonding regions separately. The electrostatic perturbation then takes the form

$$V = \sum_{ab} \sum_{tu} \hat{Q}_t^a T_{tu}^{ab} \hat{Q}_u^b, \quad (11)$$

where \hat{Q}_t^a is the operator for one of the multipole moments of region a of molecule A . The electrostatic energy becomes

$$U_{es}^{AB} = \sum_{ab} \sum_{tu} Q_t^a T_{tu}^{ab} Q_u^b. \quad (12)$$

In defining these regional moments it is necessary to specify an origin for each region, and the origin for region a is known as 'site a '. In the Distributed Multipole approach we replace the extended charge distribution of region a by a set of point multipoles at site a , rather than replacing the charge distribution of the entire molecule by a set of point multipoles at the molecular origin.

The principle of the Distributed Multipole Analysis method[10] is that the charge distribution of any molecule can be represented in the form

$$\sum_{ij} \rho_{ij} \eta_i \eta_j,$$

where η_i and η_j are Gaussian primitive orbitals of the type used in modern wave-function calculations and ρ_{ij} is a density matrix element. Now the product $\eta_i^a \eta_j^b$ of orbitals centred at positions a and b is another Gaussian function centred at a point p on the line from a to b . This product charge distribution can be represented *exactly* in the long-range limit by a terminating multipole expansion about p . For example, in the case where η_i and η_j are both p functions, the multipole expansion about p terminates at the quadrupole term. If the potential due to the charge distribution is instead represented by a multipole expansion about some other point at distance r from p , the terms in that multipole expansion behave like $(r/R)^n$, where R is the distance R at which the potential is required, so the expansion converges rapidly if r is small compared with R .

Accordingly we choose a number of sites in the molecule, usually but not necessarily at the nuclei, and represent each overlap distribution by a multipole expansion about the nearest site. This gives a description of the charge distribution that has the best possible convergence properties for the given choice of sites. The sites can be chosen to suit the problem being investigated and the accuracy required; for instance, it is often helpful in studies of small molecules to add sites at bond centres, while in larger systems groups of atoms like a methyl group can be represented by a single site. In this way we reduce the elaborate and time-consuming evaluation of the expression (5) to the very much simpler calculation of (8).

The simplest possible reliable model for the intermolecular interaction between polar molecules is based on the distributed multipole picture, though it is necessary to add some form of repulsive potential. Buckingham & Fowler added a simple hard-sphere repulsion, using standard van der Waals radii,

to obtain a model that has been extremely successful in predicting the structures of hydrogen-bonded van der Waals complexes[11]. Several authors have confirmed by using more detailed calculations that the electrostatic interaction is indeed dominant in determining the structures of such complexes[12-14]. The method has also been applied to systems involving aromatic molecules[15], where it is very successful, contradicting the common perception that because such molecules are non-polar electrostatic effects are unimportant. Of course it has to be recognised that other effects, especially dispersion, are also important in such systems.

Computation of Distributed Multipoles

The Distributed Multipole description of a molecular charge distribution is not unique; it depends on the choice of sites and on the precise definition of the region boundaries. (The conventional single-site multipole description is not unique either, since it depends on the choice of origin[16].) There are many ways of determining these distributed multipole moments[10,17-23]. For example, Lavery, Etchebest & Pullman[19] calculated the moments of localised molecular orbitals. By locating the expansion centre at the centroid of the orbital, they ingeniously eliminated the dipole contributions, but unfortunately the convergence properties of this description are poor because the localised orbitals have tails that extend across much of the molecule. Moreover this description cannot easily be applied to correlated wavefunctions. Cooper & Stutchbury[21] used Bader partitioning of the charge distribution; this has some conceptual attractions but again the convergence properties of the resulting description are poor[10]. While these and other methods have some merits, the over-riding criterion for the choice of an electrostatic model is the rate of convergence of the resulting multipole expansion, and the DMA procedure was designed to optimise this, as was the technique used by Vigné-Maeder & Claverie[23]. It is implemented as a part of the Cambridge Analytic Derivatives Package (CADPAC)[24], and is able to evaluate Distributed Multipoles up to rank 10 if required, at sites chosen by the user. The rank of multipoles on any particular site can be limited, in which case the effects of higher multipoles on that site are described by moments on neighbouring sites. For example, it is often convenient to specify that hydrogen atoms may carry a charge but no dipole or higher moment. The same procedure is also available in the Direct SCF program of Murray & Amos[25], and this makes it possible to perform calculations on molecules containing over 100 atoms, such as small polypeptides and the constituents of liquid crystals. Faerman & Price[26] have used the Direct SCF program to calculate Distributed Multipoles for a number of dipeptides and amides, and have been able to obtain a transferable multipole model for such systems, including moments up to octopole on each heavy-atom site. This transferable model gives a good account of the potential around cyclosporin, an 11-residue polypeptide, which was also studied using the Direct SCF method[27].

There are two popular methods for modelling molecular charge distributions that are best avoided. These are the use of Mulliken populations and 'potential-fitted point charges'. Mulliken populations have been widely used, but the only argument in their favour is that they are readily available because they are calculated by almost all *ab initio* programs. They are widely acknowledged to give an poor description of the molecular charge distribution, but it does not seem to be generally understood why this is so. The reason is that although they give a reasonable account of the distribution of charge between atoms, they take no account of the distortion of individual atoms. That is, they ignore atomic dipoles and quadrupoles. Since the atomic dipoles make a substantial contribution to the molecular dipole, this means that Mulliken populations give wildly inaccurate estimates of molecular dipoles and hence of the long-range electrostatic potential. Since the importance of atomic dipoles and quadrupoles is even greater at short range, the Mulliken charge distribution is useless for modelling short-range

interactions.

A better method, by comparison, is the 'potential-fitted point charge' approach[28,29]. Here the molecular electrostatic potential is calculated from an *ab initio* wavefunction, and a point-charge model is derived by fitting atomic charges so as to reproduce the potential. This gives a better description of the charge distribution than Mulliken charges, but as we have seen this is not a remarkable achievement. The method has little else to commend it. Distributed Multipole Analysis[10] has shown that atomic dipoles and quadrupoles make very important contributions to the overall molecular electrostatic potential, especially at short range; a point charge model can only mimic these effects by assigning charges to neighbouring atoms, and the resulting description is bound to have poor convergence properties at short range. Moreover the method used to compute the point charges is fairly time-consuming, and is also very indirect. Distributed Multipole Analysis, on the other hand, gives a very fast and direct route from any wavefunction to a description of the charge distribution that represents features such as lone pairs and π electrons in a straightforward and natural way, and is guaranteed to reproduce the electrostatic potential exactly in the long-range limit as well as providing a very good approximation to it at all accessible configurations.

A further disadvantage of the potential-fitted point-charge approach is that at short distances there are penetration contributions to the electrostatic potential (discussed further below) and these cannot be represented by any expansion in powers of $1/R$ because they decay exponentially with distance. Such effects should be included in the short-range part of the potential; attempts to include them in the point-charge model merely distort the model without improving its accuracy.

Perturbation Theory at Short Range

Perturbation theory becomes troublesome when the intermolecular distance is short. This is not because the perturbation becomes too large; for configurations that are accessible at thermal energies the interaction is usually still very small compared with the separation between electronic energy levels, so perturbation theory should still converge rapidly. The problem arises from the overlap of the wavefunctions, and the consequences for perturbation theory are profound. In the first place, it becomes impossible to distinguish between electrons that 'belong' to molecule *A* and electrons that 'belong' to *B*. It is therefore no longer possible to separate the Hamiltonian for the entire system in the manner of eq. (2), and there is no satisfactory way to define a perturbation operator representing the interaction between the molecules. In the second place, the wavefunctions for the unperturbed system cannot be taken to be simple products of the form $\psi_m^A \psi_n^B$; they have to be antisymmetrized with respect to all electron permutations. A more serious difficulty is that whether antisymmetrized or not, the functions $\psi_m^A \psi_n^B$ are not orthogonal to each other. Ordinary Rayleigh-Schrödinger perturbation theory assumes that the unperturbed Hamiltonian has a complete set of eigenfunctions, and that they are orthogonal.

Many versions of perturbation theory have been proposed to overcome these problems. A large number of them rely on an expansion of the perturbation equations in powers of the overlap between the functions on *A* and those on *B*. This approach appears to work when small basis sets are used, but as the basis is improved, the overlap between the functions on the two molecules becomes larger, and the expansion ceases to converge. This failure of the overlap expansion occurs with quite modest basis sets. Accordingly it is necessary to use a method that deals explicitly with the natural non-orthogonal basis functions for the problem.

If we have to use non-orthogonal wavefunctions, then the natural one-electron orbitals in which to express them are the SCF molecular orbitals of the non-interacting molecules. From these we can construct antisymmetrized (determinantal) wavefunctions in which some orbitals of each molecule are occupied. Because of the non-orthogonality of the orbitals, these determinantal wavefunctions will

also be non-orthogonal. It is possible to construct a perturbation theory in which the wavefunction is expanded in terms of these determinants. Fortunately it is possible to formulate it in such a way that the separation of the Hamiltonian into an unperturbed part and a perturbation is unnecessary. The resulting Intermolecular Perturbation Theory (IMPT)[30] has been incorporated into the Cambridge Analytical Derivatives Package (CADPAC)[24].

The first-order perturbation energy in the short-range theory can be separated into two parts: the electrostatic energy and the exchange-repulsion energy. The formal expression for the electrostatic interaction is still eq. (5) at short range, but because the charge densities overlap, it can no longer be expressed completely in terms of a multipole or distributed-multipole expansion. There are now additional terms that describe the effects arising from the interpenetration of the charge distributions. This *penetration* effect is illustrated by the simple case of a He^+ ion interacting with a proton[23,31]. The electrostatic interaction can be evaluated explicitly in this case, and is

$$U_{\text{es}}(\text{He}^+ \cdots \text{H}^+) = \frac{1}{R} + \left(Z + \frac{1}{R} \right) e^{-2ZR},$$

where R is the separation. Here we can distinguish two terms: the 'multipolar part' $1/R$ which is the classical repulsion between two unit charges at distance R , and the 'penetration term' $(Z + 1/R)\exp(-2ZR)$ (where $Z = 2$ is the helium nuclear charge) which describes the modification to the multipolar expression that arises from the penetration of the proton within the electronic charge distribution of the He^+ . We see that the latter decays exponentially with separation. It is often said that the multipole expansion of the complete potential is an *asymptotic* expansion in $1/R$, because it is impossible to find a convergent series in $1/R$ for an exponential e^{-aR} . However it is much more satisfactory to separate the multipolar part of the interaction, which converges under well-defined and easily attainable conditions[23,31], from the exponential penetration part, for which any attempt at an expansion in powers of $1/R$ is pointless, and which is much better regarded as part of the short-range interaction.

Unfortunately we have no way at present of evaluating the penetration part of the electrostatic interaction other than by taking the difference between the exact electrostatic energy (5), which includes the penetration effects, and the multipolar approximation (8), which does not. This is clearly no help at all. However Hall has found[32] that a good account of the molecular electrostatic potential, including penetration effects, is given by a 'current bun' model comprising point charges together with a sum of a small number of spherical Gaussian charge distributions. The point charges in such a model yield a form of distributed multipole expansion, while the spherical Gaussians yield both multipolar and penetration contributions. Here we are dealing with the electrostatic interaction between a molecule and a formal test charge, but a similar idea may be helpful in describing the interaction between two molecules. Whether or not this particular approach does prove fruitful, there is a good prospect that we can find an efficient way to describe the penetration effects as well as the long-range effects in terms of monomer properties.

The second additional term that appears in the first-order energy at short range is the exchange-repulsion energy, which is much more difficult to deal with. It comprises two effects. an attractive exchange term which arises because the electrons can exchange between the molecules, and a repulsive term which occurs because the electrons cannot occupy the same region of space if they have the same spin (Pauli repulsion). However it is not usually helpful to separate these two parts. Because the exchange-repulsion is a first-order perturbation term, it is still only necessary to know the unperturbed wavefunctions to evaluate it. Nevertheless it is time-consuming to calculate, because all the intermolecular electron-repulsion integrals are needed, and they are different for each configuration of the complex. Moreover many of the usual simplifications that arise in the evaluation of two-electron

integrals do not apply, because the wavefunctions of the two molecules are not orthogonal. The labour is even greater if correlated wavefunctions are used for the monomers, though Rijks *et al.*[33] have suggested how the computational effort may be reduced.

Unfortunately there is as yet no known way to obtain the repulsion energy from properties of the separate molecules. An attempt has been made to characterise the repulsive surface of a molecule by performing IMPT calculations between the molecule and a suitable test particle, such as a helium atom. Because the helium atom has only one molecular orbital and is spherically symmetrical, such calculations can be done much more easily than calculations involving two ordinary molecules. From the data for the repulsion between molecule *A* and the test particle, and between *B* and the test particle, it may be possible to construct a repulsive potential between *A* and *B*. Some limited progress has been made with this idea[34]. An alternative approach[35] has been based on the suggestion[36-38] that the repulsion energy is closely correlated with the overlap between the molecular wavefunctions, but this seems likely to be more useful as a guide to the form of analytic models than as a direct route to accurate potential functions.

Consequently it is necessary at present to rely on suitable models of the short-range contributions to the energy. A successful form of model repulsion potential is the following *anisotropic atom-atom repulsion*:

$$U_{\text{rep}} = \sum_{ab} \exp[\alpha(\Omega_{ab})(R_{ab} - \rho(\Omega_{ab}))], \quad (13)$$

in which the shape parameter ρ_{ab} and the hardness parameter α_{ab} depend on the relative orientation Ω_{ab} of atoms *a* and *b*. (In the general case, Ω_{ab} is a short-hand for five orientation variables.) There is now overwhelming evidence[31] of the need to introduce this kind of anisotropy into atom-atom repulsions. More elaborate forms for the radial dependence have been used for very accurate potentials, for example to describe the Ar...Ar interaction[39], but there is not enough information at present to characterise the repulsion in this much detail for larger systems. There it is necessary to fit a potential model such as (13) to whatever data may be available, either experimental data such as crystal structures or calculated potentials for smaller systems.

One of the attractive features of the repulsion model (13) is that because of its intrinsic anisotropy it provides a natural way to describe repulsions between small groups of atoms, so that it is not necessary to treat every atom individually. In the azabenzenes, for instance, which are derived from benzene by replacing one or more CH units by N atoms, it was possible to obtain a model of the intermolecular potential in which each CH group and each N atom was treated as a single unit. The CH...CH, CH...N and N...N repulsive potentials were transferable between different molecules, and the multipole moments were also transferable except that the short-range inductive effect was described by the transfer of approximately $\frac{1}{4}$ of an electron from C to N along every CN bond. Empirical dispersion terms were also included. The resulting model was very successful in describing the very varied crystal structures of a number of azabenzenes[40].

A similar approach has been very successful in studies of the halogens, Cl₂, Br₂ and I₂. These molecules all have *Cmca* crystal structures, unlike most diatomics, and this cannot be understood until it is appreciated that the atoms in these molecules are significantly non-spherical, being flattened at the ends, by as much as 20% in the case of I₂. The crystal structures are then easily understood as a consequence primarily of packing effects[41]. A distributed-multipole description of the electrostatics and an anisotropic atom-atom dispersion term were added to complete the potential model, which was then able to give a very good account of the liquid structure and properties and the lattice frequencies, despite being fitted only to the crystal structure and lattice energy[42].

Long-range perturbation theory—second-order effects

A similar philosophy can be applied to the higher-order terms in the perturbation expansion. The second-order energy is, according to Rayleigh-Schrödinger perturbation theory,

$$W^{(2)} = - \sum'_{mn} \frac{\langle 00|V|mn\rangle \langle mn|V|00\rangle}{W_m^A + W_n^B - W_0^A - W_0^B}. \quad (14)$$

This is conventionally separated into three terms: those in which $n \neq 0$ but $m = 0$, those in which $m \neq 0$ but $n = 0$, and those in which neither m nor n is 0. These three terms are

$$U_{\text{ind}}^A = - \sum'_m \frac{\langle 00|V|m0\rangle \langle m0|V|00\rangle}{W_m^A - W_0^A}, \quad (15)$$

$$U_{\text{ind}}^B = - \sum'_n \frac{\langle 00|V|0n\rangle \langle 0n|V|00\rangle}{W_n^B - W_0^B}, \quad (16)$$

$$U_{\text{disp}} = - \sum'_m \sum'_n \frac{\langle 00|V|mn\rangle \langle mn|V|00\rangle}{W_m^A + W_n^B - W_0^A - W_0^B}. \quad (17)$$

In the first of these, we may perform the integrations over the ground-state wavefunction of B to obtain

$$U_{\text{ind}}^A = - \sum'_m \frac{\langle 0|V^B|m\rangle \langle m|V^B|0\rangle}{W_m^A - W_0^A}, \quad (18)$$

where $V^B = \langle \psi_0^B|V|\psi_0^B\rangle$ is the potential at A due to the unperturbed charge distribution of B . Eq. (18) then describes the response of molecule A to this potential, and is the induction energy of molecule A in the field of B . Similarly eq. (16) is the induction energy of B in the field of A .

These induction energy expressions can be reformulated so as to depend only on properties of the individual molecules. To do this, we start from the expression (18) for the induction energy of molecule A , and use the expression (6) for the perturbation V . We evaluate the integral over the coordinates of molecule B , and arrive at the expression

$$\begin{aligned} U_{\text{ind}}^A &= - \sum'_m \sum_{aa'bb'} \sum_{tt'uu'} Q_u^b T_{ut}^{ba} \frac{\langle 0|\hat{Q}_t^a|m\rangle \langle m|\hat{Q}_{t'}^{a'}|0\rangle}{W_m^A - W_0^A} T_{t'u'}^{a'b'} Q_{u'}^{b'} \\ &= - \frac{1}{2} \sum_{aa'bb'} \sum_{tt'uu'} Q_u^b T_{ut}^{ba} \alpha_{tt'}^{aa'} T_{t'u'}^{a'b'} Q_{u'}^{b'}, \end{aligned} \quad (19)$$

where

$$\alpha_{tt'}^{aa'} = \sum'_m \frac{\langle 0|\hat{Q}_t^a|m\rangle \langle m|\hat{Q}_{t'}^{a'}|0\rangle + \langle 0|\hat{Q}_{t'}^{a'}|m\rangle \langle m|\hat{Q}_t^a|0\rangle}{W_m^A - W_0^A}. \quad (20)$$

Here $\alpha_{tt'}^{aa'}$ is a polarizability that describes the response of the moment t at site a to a perturbation (a change in potential, field, field gradient, etc.) at site a' [43].

Now the expression (20) is an uncoupled formulation of the polarizability. We can replace it by a polarizability derived from coupled Hartree-Fock perturbation theory, which is more accurate, because it takes account of the reorganisation of the electron distribution in a self-consistent manner. Better still would be to evaluate the monomer polarizability by a method that takes account of electron correlation as well[44]. But whatever the level of calculation, we can once again perform a much better calculation of the monomer property than is possible for the dimer. In this way we arrive at a description of the induction energy that is far more accurate than we can obtain through either intermolecular perturbation theory, where the perturbation is treated in an uncoupled fashion, or from

a supermolecule calculation, where the size of the basis is limited by the need to perform calculations at a large number of points on the potential energy surface.

Distributed Polarizabilities

The distributed polarizabilities $\alpha_{i''}^{aa'}$ that occur in eq. (19) for the induction energy may be unfamiliar objects. They describe the change in the charge distribution of region a of the molecule (in terms of changes to the charge, dipole moment, and so forth) that are caused by a change in the electrostatic fields (potential, electric field, field gradient, etc.) at site a' .

If sites a and a' are the same, then we have a *local* polarizability. For example, the polarizability $\alpha_{10,10}^{aa}$ describes the change in the z component of the dipole of region a that results from an electric field in the z direction at site a . (Remember that the moment Q_{10} is the same as the dipole component μ_z ; see eq. (10).) This is an ordinary local dipole polarizability. However the induced dipole of region a changes the field that is experienced by the other sites, so there are other non-local effects of the original field at a , and these are described by the non-local polarizabilities $\alpha_{i''}^{aa'}$ in which $a \neq a'$. Because they describe secondary polarizations they are usually smaller than the local polarizabilities.

When a molecule is divided into several regions, it becomes possible for charge to flow between the regions in response to a potential difference between them. This means that there are charge-flow polarizabilities $\alpha_{00,00}^{aa'}$ (remember that multipole Q_{00} is the charge) that describe the change in the charge of region a when the potential at site a' is changed. The total charge on the molecule must remain constant, so the total change in charge must be zero; this is expressed by a sum-rule for the charge-flow polarizabilities:

$$\sum_a \alpha_{00,00}^{aa'} = 0.$$

It follows that the non-local charge-flow polarizabilities are of similar magnitude to the local ones; if the charge in region a changes by a certain amount, the charges on the neighbouring sites must change by amounts of similar magnitude for the total charge to remain constant.

The reason for using distributed polarizabilities to describe the distortion of the charge distribution is the same as the reason for using distributed multipoles to describe the charge distribution itself. It is a matter of convergence and efficiency of description. The electrostatic field at molecule A of a neighbouring molecule B is strongly non-uniform, and a Taylor expansion about the centre of A will converge very slowly, if at all, at points far removed from that centre. If it does converge, an accurate description of the field requires many terms in the Taylor expansion, for which high-order derivatives of the potential are needed. The effect of these high-order derivatives of the field on the charge distribution of A is described by high-rank polarizabilities. If the Taylor expansion of the electrostatic potential of B does not converge at some points within A , then no single-centre polarizability description of the distortion is possible. If on the other hand, we study each small region of A separately, the convergence is much better, and the description of the distortion can be achieved by using distributed polarizabilities of relatively low rank. The inadequacy of the single-centre description is illustrated by the observation[45] that even with distributed polarizabilities it is still necessary to include terms up to quadrupole rank at least to obtain a satisfactory description of the induction energy.

Distributed polarizabilities are also needed for calculating the induced moments of van der Waals complexes. For the reasons outlined above, single-centre polarizabilities are not satisfactory for calculating the moments induced in one molecule of a complex by its neighbours, and some apparent anomalies were found when such calculations were attempted. Induced moments calculated using distributed polarizabilities show much more satisfactory agreement with experiment, and the apparent anomalies can be understood in terms of features of the distributed multipoles or distributed polar-

izabilities of the molecules concerned[46]. A simple example is the case of the complexes $N_2 \cdots HCl$ and $OC \cdots HCl$, which have rather different induced moments although the polarizabilities of N_2 and CO are almost identical[47]. The distributed polarizabilities show that the C atom of CO is more polarizable, and the O atom less polarizable, than the N atom in N_2 . Since the C is closer to the HCl , it is in a stronger field than is the O, and the overall induced dipole is larger. It is also the case that the atomic dipole on the C atom in CO is larger than that on the N atoms in N_2 , so that the HCl is more strongly polarized by the CO than by the N_2 [46].

There is another way in which the expression (19) might be improved upon. It gives the induction energy of molecule A in the field arising from the multipole moments of molecule B . However molecule B is also polarizable, and its moments will be modified by the presence of molecule A . If this effect is taken into account, we arrive at an expression for the induction energy that is a power series in the molecular polarizabilities[48]. In practice, the effects of molecular polarization are usually calculated in an iterative fashion; the polarized moments of each molecule are evaluated in the field due to the other molecules, and the calculation is repeated until the polarized moments are self-consistent. This however is equivalent[48] to taking some, but not all, of the terms in a perturbation series to infinite order, and moreover it is known[49] that the perturbation series for the induction energy is asymptotic, i.e. divergent. This means that the conventional iterative procedure is highly questionable, and indeed it is known to lead to singularities at short range[50]. Numerical investigation[45] suggests that the simple expression (19) is more satisfactory, provided that distributed polarizabilities are used and provided that polarizabilities up to at least quadrupole rank are included.

As in the case of the first-order energy, there are two modifications that have to be made to this description when the wavefunctions overlap. In the first place, there must be a penetration effect that arises from the overlap of the charge densities. This has not been examined in detail, but it appears to be small[45]. The second modification arises because when two molecules overlap, it becomes possible for electron density from either molecule to flow onto the other. This effect is called charge transfer. In perturbation theory terms, it can be described by excitations from the occupied orbitals of one molecule to the virtual orbitals of the other. As such, it incorporates not only the genuine physical effect of charge transfer, but the BSSE. One of the virtues of IMPT is that the effects of BSSE do not arise in most of the energy terms, and the charge-transfer interaction is the only term involving single excitations in which they do occur. It is possible to correct the calculated charge-transfer energy for these effects by a procedure similar to that used in the supermolecule method, but as in that case some uncertainties remain.

There is a further problem with charge transfer, however. If we could perform the calculation with a complete basis set on molecule A , then it would be possible to describe any virtual orbital of B in terms of the basis set for A . In this case, the charge transfer effects would be included completely in the induction energy for molecule A . If we were then to calculate the effects of excitations from occupied orbitals of A to virtual orbitals of B we would be counting the same effects again. In other words, the charge-transfer energy is formally spurious. In practice, the basis sets used today are too small for this to happen to any great extent, but we should be aware that in principle the charge-transfer contribution is subject to this kind of double-counting error.

The dispersion energy

The expression (17) for the dispersion energy can also be replaced by a more accurate expression in terms of monomer properties. Eq. (17) involves one-electron excitations on both molecules, and cannot appear in an SCF calculation on the supersystem; accordingly the dispersion energy is a manifestation of electron correlation. Nevertheless, because there is a single excitation on each molecule, the

dispersion energy can be reformulated in terms of monomer polarizabilities, which are one-electron properties and can be calculated reasonably accurately at the SCF level.

The key is the replacement of the energy denominator via the Casimir-Polder identity[51], which can be established by a simple contour integration:

$$\frac{1}{A+B} = \frac{2}{\pi} \int_0^\infty \frac{AB}{(A^2+u^2)(B^2+u^2)} du, \quad \text{for } A > 0, B > 0. \quad (21)$$

In eq. (17) we put $A = W_m^A - W_0^A$ and $B = W_m^B - W_0^B$, and then the use of eq. (21), together with the expression (6) for the perturbation, leads to the following expression for the dispersion energy:

$$U_{\text{disp}} = -\frac{\hbar}{2\pi} T_{tu}^{ab} T_{t'u'}^{a'b'} \int_0^\infty \alpha_{tt'}^{aa'}(iu) \alpha_{uu'}^{bb'}(iu) du, \quad (22)$$

where the polarizability at imaginary frequency $\alpha_{tt'}^{aa'}(iu)$ is given by

$$\alpha_{tt'}^{aa'}(iu) = \sum_m \frac{\Delta W_m (\langle 0 | \hat{Q}_t^a | m \rangle \langle m | \hat{Q}_{t'}^{a'} | 0 \rangle + \langle 0 | \hat{Q}_{t'}^{a'} | m \rangle \langle m | \hat{Q}_t^a | 0 \rangle)}{(\Delta W_m)^2 + \hbar^2 u^2}, \quad (23)$$

with $\Delta W_m = W_m - W_0$. This is a much more tractable and useful formulation than it appears to be at first sight. Once again the multipole moment operators \hat{Q}_t^a are referred to *local* axes in the molecule, so the polarizabilities defined by (23) are also referred to local axes. Consequently no information about the relative orientation of the molecules is required to evaluate the dispersion integrals in (22), and they can be obtained once and for all for any pair of molecules, using accurate coupled Hartree-Fock calculations on the monomer to obtain the polarizabilities. The need to evaluate polarizabilities at imaginary frequency is not a problem; they can be calculated just as easily as the static polarizabilities. Moreover eq. (23) shows that $\alpha_{tt'}^{aa'}(iu)$ is very well-behaved as a function of frequency (it tends monotonically to zero as $u \rightarrow \infty$) so the dispersion integrals can be evaluated accurately by numerical quadrature using the values of $\alpha_{tt'}^{aa'}(iu)$ at a dozen or so frequencies[44].

The expression (22) for the dispersion energy is rather cumbersome, since it involves non-local polarizabilities $\alpha_{tt'}^{aa'}(iu)$. Many of these are small, and it is possible to transform the expression for the dispersion energy in such a way that their effects are represented exactly, at sufficiently large distances, by local terms[52]. However this cannot always be done for the dispersion integrals involving charge-flow polarizabilities on one or both molecules, and these dispersion integrals contribute significantly to the overall dispersion energy. It seems likely that the conventional site-site picture of the dispersion interaction is invalid, or at least incomplete, for large conjugated molecules.

It should eventually be possible to obtain accurate polarizabilities and dispersion integrals using correlated wavefunctions. Unfortunately there are some difficulties[53], but these are being overcome[44].

Once again there are corrections to be made in the short-range region. Since the dispersion energy is part of the correlation energy for the supersystem, it must remain finite at short range, while the terms in the multipole expansion (22) diverge like some power of $1/R$. It is usual to multiply the dispersion expression by a 'damping function' to cancel this singularity. Several authors have suggested suitable damping functions[54-56].

SUMMARY

The calculation of accurate intermolecular potentials is a difficult task. Supermolecule calculations are expensive, and are subject to inaccuracies that are difficult to avoid or correct. An approach based on

perturbation theory principles holds some promise. The intermolecular potential is obtained as a sum of several terms, many of which can be calculated in terms of properties of the separate molecules, which can be obtained much more accurately than is possible when calculations have to be carried out on the whole complex. We are led in this way to seek reliable descriptions of the charge distribution and polarizability of the individual molecules. From these it is possible to construct accurate descriptions of the electrostatic, induction and dispersion energy, leaving only the short-range terms in the potential energy to be modelled by more empirical methods.

The *electrostatic energy* is obtained in terms of distributed multipoles, and is corrected at short range for the effects of penetration. The multipole moments can be calculated straightforwardly from accurate wavefunctions for the individual molecules. It is necessary to use multipole moments at least as high as quadrupole if multipole sites are taken on each atom. This means that the expression for the electrostatic energy is more complicated than for point charges, but it has been shown that it can be handled efficiently even in the demanding circumstances of molecular dynamics calculations[42], and the necessary formulae are all in the literature[9,7].

The *induction energy* is expressed in terms of distributed multipoles and distributed polarizabilities. It is more difficult to handle than the electrostatic interaction, but the same interaction functions occur. The *dispersion energy* can be expressed in terms of polarizabilities at imaginary frequency for the individual molecules, and the use of distributed polarizabilities makes it possible to describe the dispersion interaction accurately at short range. A certain amount of manipulation is needed to transform the energy expression into manageable form, but when this is done the resulting picture is very similar to the conventional atom-atom dispersion models that are widely used, except that it includes additional terms that arise from charge-flow polarizabilities in the distributed picture. Damping functions must be included to correct for the effects of overlap[54-56]. The effects of electron correlation should be included in the polarizabilities; this is currently difficult to do, but the principles are understood[44].

The techniques are therefore all available for a new generation of potentials for intermolecular and intramolecular interactions. With their help it should be possible to model such interactions in biological systems much more accurately than has previously been possible.

References

- [1] S. L. Price & A. J. Stone, *Molec. Phys.* 40 (1980) 805.
- [2] R. J. Bartlett, *J. Phys. Chem* 93 (1989) 1697.
- [3] J. A. Pople, M. Head-Gordon & K. Raghavachari, *J. Chem. Phys.* 87 (1988) 5968.
- [4] S. F. Boys & F. Bernardi, *Molec. Phys.* 19 (1970) 553.
- [5] J. H. van Lenthe, J. G. C. M. van Duijneveldt-van der Rijdt & F. B. van Duijneveldt, *Adv. Chem. Phys.* 69 (1987) 521.
- [6] K. Szalewicz, S. J. Cole, W. Kolos & R. J. Bartlett, *J. Chem. Phys.* 89 (1988) 3662.
- [7] A. J. Stone, in *Theoretical Models of Chemical Bonding*, vol. 4, Z. B. Maksić, ed., Springer (1989).
- [8] A. J. Stone & R. J. A. Tough, *Chem. Phys. Lett.* 110 (1984) 123.
- [9] S. L. Price, A. J. Stone & M. Alderton, *Molec. Phys.* 52 (1984) 987.
- [10] A. J. Stone, *Chem. Phys. Lett.* 83 (1981) 233; A. J. Stone & M. Alderton, *Molec. Phys.* 56 (1985) 1047.
- [11] A. D. Buckingham & P. W. Fowler, *J. Chem. Phys.* 79 (1983) 6426; *Canad. J. Chem.* 63 (1985) 2018.
- [12] G. J. B. Hurst, P. W. Fowler, A. J. Stone & A. D. Buckingham, *Int. J. Quantum Chem.* 29 (1986) 1223-1239.
- [13] A. P. L. Rendell, G. B. Bacskay & N. S. Hush, *Chem. Phys. Letters* 117 (1985) 400-408.

- [14] V. Magnasco, C. Costa & G. Figari, *Chem. Phys. Lett.* (1989) .
- [15] S. L. Price & A. J. Stone, *J. Chem. Phys.* **86** (1987) 2859.
- [16] A. D. Buckingham, *Adv. Chem. Phys.* **12** (1967) 107.
- [17] M. Mezei & E. S. Campbell, *Theor. Chim. Acta* **43** (1977) 227.
- [18] A. Pullman & D. Perahia, *Theor. Chim. Acta* **48** (1978) 29.
- [19] R. Lavery, C. Etchebest & A. Pullman, *Chem. Phys. Lett.* **85** (1982) 266.
- [20] J. F. Rico, J. R. Alvarez-Collado & M. Paniagua, *Molec. Phys.* **56** (1985) 1145.
- [21] D. L. Cooper & N. C. J. Stutchbury, *Chem. Phys. Lett.* **120** (1985) 167.
- [22] W. A. Sokalski & A. Sawaryn, *J. Chem. Phys.* **87** (1987) 526.
- [23] F. Vigné-Maeder & P. Claverie, *J. Chem. Phys.* **88** (1988) 4934.
- [24] R. D. Amos & J. E. Rice, CADPAC: The Cambridge Analytical Derivatives Package, issue 4.0, Cambridge, 1987.
- [25] C. W. Murray & R. D. Amos, in preparation.
- [26] C. F. Faerman & S. L. Price, *J. Amer. Chem. Soc.*, submitted for publication.
- [27] S. L. Price, R. J. Harrison & M. F. Guest, *J. Comput. Chem.* **10** (1989) 552.
- [28] F. A. Momany, *J. Phys. Chem* **82** (1978) 592.
- [29] U. C. Singh & P. A. Kollman, *J. Comput. Chem.* **5** (1984) 129.
- [30] I. C. Hayes & A. J. Stone, *Molec. Phys.* **53** (1984) 69; I. C. Hayes & A. J. Stone, *Molec. Phys.* **53** (1984) 83; G. J. B. Hurst, I. C. Hayes & A. J. Stone, *Molec. Phys.* **53** (1984) 107.
- [31] A. J. Stone & S. L. Price, *J. Phys. Chem* **92** (1988) 3325.
- [32] G. G. Hall & K. Tsujinaga, *Theor. Chim. Acta* **69** (1986) 425; K. Tsujinaga & G. G. Hall, *Theor. Chim. Acta* **70** (1986) 257.
- [33] W. Rijks, M. Gerritsen & P. E. S. Wormer, *Molec. Phys.* **66** (1989) 929.
- [34] A. J. Stone & C.-S. Tong, in preparation.
- [35] R. J. Wheatley & S. L. Price, , submitted for publication.
- [36] S. Kita, K. Noda & H. Inouye, *J. Chem. Phys.* **64** (1976) 3346.
- [37] Y. S. Kim, S. K. Kim & W. D. Lee, *Chem. Phys. Lett.* **80** (1981) 574.
- [38] P. Gellert, D.Phil. thesis, University of Oxford.
- [39] G. C. Maitland, M. Rigby, E. B. Smith & W. A. Wakeham, *Intermolecular Forces*, Clarendon Press, Oxford (1981).
- [40] S. L. Price & A. J. Stone, *Molec. Phys.* **51** (1984) 569.
- [41] S. L. Price & A. J. Stone, *Molec. Phys.* **47** (1982) 1457.
- [42] P. M. Rodger, A. J. Stone & D. J. Tildesley, *J. Chem. Soc. Faraday II* **83** (1987) 1689; P. M. Rodger, A. J. Stone & D. J. Tildesley, *Molec. Phys.* **63** (1988) 173; P. M. Rodger, A. J. Stone & D. J. Tildesley, *Chem. Phys. Letters* **145** (1988) 365.
- [43] A. J. Stone, *Molec. Phys.* **56** (1985) 1065.
- [44] P. E. S. Wormer & W. Rijks, *Phys. Rev. A* **33** (1986) 2928; W. Rijks & P. E. S. Wormer, *J. Chem. Phys.* **88** (1988) 5704.
- [45] A. J. Stone, *Chem. Phys. Lett.* **155** (1989) 111.
- [46] A. D. Buckingham, P. W. Fowler & A. J. Stone, *Internat. Rev. Phys. Chem.* **5** (1986) 107; P. W. Fowler & A. J. Stone, *J. Phys. Chem* **91** (1987) 509.
- [47] R. S. Altman, M. D. Marshall & W. Klemperer, *J. Chem. Phys.* **79** (1983) 52; R. S. Altman, M. D. Marshall & W. Klemperer, *J. Chem. Phys.* **79** (1983) 57.
- [48] A. J. Stone, *Chem. Phys. Lett.* **155** (1989) 102.
- [49] A. Dalgarno & A. L. Stewart, *Proc. Roy. Soc. A* **238** (1956) 276; A. Dalgarno & N. Lynn, *Proc. Phys. Soc. London, A* **70** (1957) 223.
- [50] A. D. Buckingham & J. A. Pople, *Trans. Faraday Soc.* **51** (1955) 1173.
- [51] H. B. G. Casimir & D. Polder, *Phys. Rev.* **73** (1948) 360.
- [52] A. J. Stone & C.-S. Tong, *Chem. Phys.*, in press.
- [53] F. Visser, P. E. S. Wormer & P. Stam, *J. Chem. Phys.* **79** (1983) 4973; F. Visser, P. E. S. Wormer & W. P. J. H. Jacobs, *J. Chem. Phys.* **82** (1984) 3753; F. Visser & P. E. S. Wormer, *Molec. Phys.* **52** (1984) 723.
- [54] C. Douketis, G. Scoles, S. Marchetti & A. J. Thakkar, *J. Chem. Phys.* **76** (1982) 3057.
- [55] K. T. Tang & J. P. Toennies, *Chem. Phys.* **80** (1984) 3276.
- [56] P. J. Knowles & W. J. Meath, *Chem. Phys. Lett.* **124** (1986) 164; *Molec. Phys.* **59** (1986) 965; *Molec. Phys.* **60** (1987) 1143.

DISCUSSION

BUCKINGHAM - You referred to the need for damping of the long-range interactions in the region of overlap of the electron distributions of the monomers. Can one avoid this by a suitable choice of the repulsive potential, thereby reducing the number of adjustable parameters.

STONE - If the dispersion damping were to be described by adding terms to the repulsive potential, they would have to behave like R^{-6} , R^{-8} , ... at small R to cancel the singularities in the long-range dispersion energy. The repulsion energy itself does not contain terms of this sort. If the separation R does not become too small, the effects of damping could probably be described by modifications to the repulsive potential. However the use of multiplicative damping factors is more satisfactory, and the form of damping function has been investigated in some detail (see ref. 55 of my paper).

RIVAIL - You mentioned charge flow as an important contribution to the molecular polarization phenomenon. This raises a very difficult problem when one tries to define transferable properties because the molecule's electric neutrality may be violated. Do you think that there is a hope to reasonably account for this effect through phenomenological dipole or multipole distributed polarizabilities ?

STONE - In saturated systems, the flow of charge is likely to be a short-range effect, and the flow of charge along a bond can be represented by an induced dipole (and higher moments) in the bond. In conjugated systems, however, charge-flow effects extend over greater distances and cannot so easily be represented by local polarizabilities.

GRESH - About the charge-transfer term : using the Murrell-Randic-Williams perturbation treatment, this term appears as a distinct term in its own right. Using ab-initio super molecule computations, its magnitude was often overestimated because of basis set extension effects. Nevertheless with large basis sets, and particularly in studies of cation-ligand interactions, this term comes out from energy decomposition computations, with radial and angular dependences of its own. Incorporating explicitly E_{CT} in our S/BFA procedure by means of a development into

analytical formulas based on the MRW formula, we were able to account for such dependencies. I think that E_{CT} ought to be implemented in an additive systematics, with care concerning its calibration.

About the short-range repulsion term : the most common implementation of E_{rep} is under the form of atom-atom terms. It was proposed by Murrell that E_{rep} is proportional to the square of the overlap integral involving the localized orbitals. Representing E_{rep} under the form of bond-bond, bond-lone pair and lone pair-lone pair interactions, we could reproduce quite well the radial and angular evolution of the corresponding term in the first-order term of the ab-initio supermolecule computation.

STONE - In the limit of large basis sets, the genuine physical phenomenon of charge transfer is completely described as a part of the induction energy. In the sum-over-states formalism, charge-transfer is described by excitations from occupied orbitals of one molecule (A, say) to virtual orbitals of the other (B), while induction energy is described by excitations from occupied orbitals of A to virtual orbitals of A. If we use a complete basis for A, any virtual orbital of B can be expanded in terms of orbitals of A, and the $A \rightarrow B^*$ excitations duplicate $A \rightarrow A^*$ excitations. The 'charge transfer' energy described by the $A \rightarrow B^*$ excitations is therefore spurious. In a smaller basis, the duplication is less of a problem, but then the 'charge-transfer' terms are heavily contaminated by basis set superposition error. It follows that the inclusion of 'charge transfer' in $A \rightarrow B^*$ excitations is a very treacherous procedure. In any case, it has become clear in recent years that many phenomena attributed to charge transfer effects are well accounted for by electrostatic interactions, provided that the electrostatic effects are accurately described. The relationship between the overlap and the repulsion energy has been noticed by several workers. However no simple formula representing repulsion in terms of overlap has been derived from first principles, nor is it likely to exist. Moreover the overlap formulation is inadequate as an accurate route to the interaction energy, and is too complicated to use as part of a simple model.

SMITH - In your dipeptide calculations you noticed an asymmetry in the methyl group(s). This interesting phenomenon is seen in many classes of organic molecules. Your coworkers attributed it to polarisation effects. What was the evidence for this ?

STONE - The asymmetry noticed by Faerman & Price is an asymmetry of the charge distribution, rather than the geometrical asymmetry that Dr. Smith has in mind. For instance, in studies where the methyl hydrogen atoms were allowed to carry charges, those charges were not equal. It is thought that the inequalities are a consequence of

polarization in the electric fields of neighbouring atoms, but at present this is only a conjecture.

PEPE - The long range interactions play an important role in large systems, do you think that your method is accurate enough to get good results on large systems ?

STONE - Repulsions decay like $e^{-\alpha R}$ and dispersion like R^{-6} , but electrostatic interaction between neutral polar molecules behave like R^{-3} and interactions between charges like R^{-1} . Consequently it is essential to use accurate charge distributions to get the long-range interactions right.

PERAHIA - How good is the transferability of multipoles from one conformation to another ?

STONE - This is discussed in detail in the paper by Faerman & Price (ref. 26 of my paper).

ANGYAN - The idea of introducing a molecular "capacitance matrix" to describe intramolecular charge-flow effects seems to be very promising. Can you illustrate these charge-charge polarizability matrices for the case of simple polyatomic molecules (like H_2O) or for aromatic systems ?

STONE - We have some results for benzene and naphthalene, but they are not yet ready for publication.

A VIBRATIONAL MOLECULAR FORCE FIELD OF COMPOUNDS WITH BIOLOGICAL INTEREST.

MANUEL DAUCHEZ , PHILIPPE DERREUMAUX , PHILIPPE LAGANT and GERARD VERGOTEN.

Université de Lille II , Faculté de Pharmacie
Laboratoire de Génie Biologique et Médical
INSERM U279 Lille France
and
Université de Lille II , Faculté de Médecine
CERIM Lille France

SUMMARY

It is presented how an accurate and reliable molecular force field for proteins , nucleid acids and saccharides can be deduced from vibrational spectroscopies. A critical comparison is made with the more commonly used force fields .

INTRODUCTION

The knowledge of the three dimensional structure of molecules and macromolecules is of primordial importance. The molecular modeling techniques take a great part in chemistry- based disciplines such as crystallography, organic chemistry, inorganic chemistry, theoretical chemistry, polymer chemistry, medicinal chemistry, biochemistry, spectroscopy, enzymology, pharmacology and so on. There exists also a number of experimental techniques that give three dimensional pictures of molecules among which crystal structure analysis is the one that provides the most accurate structural information about biomolecules.

Computational chemistry has made larger advances since the advent of modern digital computers. Color graphics computer technology is making it increasingly possible to visualize and manipulate molecules with thousands of atoms. The computational methods leading to molecular structures, energies and properties (namely empirical, semi-empirical and *ab initio* calculations) are based on the evaluation of the molecular potential energy surface. Quantum mechanical calculations have the capabilities for being the most accurate technique but they require a great deal of computer time. Molecular mechanics use an analytical expression for the potential energy function based on empirical force constants or parameters. Once the potential energy function is known , the equilibrium structure of a molecule is obtained through minimization techniques. Molecular dynamics may also be performed in order to let the molecules reach the global

minimum energy conformation instead of a local one as obtained through energy minimization procedures and to predict their dynamical behaviour.

One of the major limitations of molecular mechanics and molecular dynamics techniques is the reliability of the force field itself. It is the aim of this paper to show how accurate force field may be determined from vibrational spectroscopy in order to improve the predictive capabilities of the molecular modeling methods.

THE MOLECULAR FORCE FIELD: FROM THE UREY-BRADLEY-SHIMANOUCHI MODEL TO THE LOCAL SYMMETRY FORCE FIELD.

A molecule is an ensemble of atomic nuclei with surrounding electrons. This ensemble is interacting with other molecules through attractive and repulsive phenomena. In the Born Oppenheimer approximation electrons are making effective responses to variations in nuclear positions. In a stable electronic state the potential energy V of a molecule is expressed as a function of nuclear positions: $V = V(x_1, y_1, z_1; x_2, y_2, z_2; \dots; x_n, y_n, z_n)$ where x_i, y_i and z_i are the cartesian coordinates of the i th atom and n the number of nuclei in the molecule.

The analytical expression of the potential energy V of a n atomic molecule is not known. The only certainty is that its value only depends on the relative position of the nuclei. It does not change with the rotation nor the translation of the molecule as a whole. For small displacements, the potential energy function is usually expanded in a Taylor series in terms of p variables q_p where p is the degree of internal freedom:

$$V = V_0 + \sum_i f_i q_i + \frac{1}{2!} \sum_{ij} f_{ij} q_i q_j + \frac{1}{3!} \sum_{ijk} f_{ijk} q_i q_j q_k + \dots$$

$$f_i = \left(\frac{\delta V}{\delta q_i} \right)_0, \quad f_{ij} = \left(\frac{\delta^2 V}{\delta q_i \delta q_j} \right)_0 \quad \text{and} \quad f_{ijk} = \left(\frac{\delta^3 V}{\delta q_i \delta q_j \delta q_k} \right)_0$$

are called the linear, quadratic and cubic force constants, respectively. As we are interested in variations of the potential energy with the displacement coordinates, its origin may be chosen as $V_0 = 0$. Then the equilibrium configuration of the molecule must be at a minimum of the potential energy and requires that all the q_i 's are independent and the linear force constants f_i are equal to zero. Again for small displacements of the nuclei around their equilibrium position, we can neglect the cubic and higher terms in the Taylor expression of V so that:

$$V = \frac{1}{2} \sum_{ij} f_{ij} q_i q_j$$

The set of f_{ij} matrix elements constitutes the General force field. The more commonly used molecular mechanics computer programs are based on a potential energy function expressed in terms of internal coordinates (bond stretching, bond angle bending and torsional coordinates) and deal with a number of internal coordinates greater than the degree of vibrational freedom $p = 3n - 6$. The problem of how to choose these p basis coordinates has been developed in detail

by T. Shimanouchi (ref.1). A number of model force fields has been presented in the literature among which the Urey-Bradley-Shimanouchi force field (UBSFF) has been very useful in explaining the vibrational spectra of many complex molecules. In this model it is assumed that the off-diagonal elements of the F matrix (composed of the f_{ij} elements) are due to the interaction between nonbonded atoms.

In describing the potential energy function , the fundamental differences with the more commonly expressions (e.g. those of CHARMM (ref.2) or AMBER (ref.3) computer programs) are highlighted. The Urey-Bradley-Shimanouchi potential energy function is composed of terms representing variations of bond stretchings , bond angle bendings, out-of-plane bendings and dihedral torsions to which Van der Waals interactions, electrostatic interactions, hydrogen-bonds are added.

$$\begin{aligned}
 V(r) = & 1/2 \sum_{\text{bonds}} K_b (r_{ij} - r_0)^2 + 1/2 \sum_{\text{angles}} H_{\theta} r_{ij} r_{jk} (\theta_{ijk} - \theta_0)^2 + 1/2 \sum_{\text{opb}} \gamma_{ijkl} (\Delta\pi_{ijkl})^2 \\
 & + 1/2 \sum_{\text{torsions}} Y_{ij} (\Delta\tau_{ij})^2 + \sum_{\substack{\text{nonbonded} \\ \text{pairs } r < 6 \text{ \AA}}} A \exp(-Br) - Cr^6 + \sum_{\substack{\text{nonbonded} \\ \text{pairs } r < 6 \text{ \AA}}} \frac{q_i q_j}{r} \\
 & + 1/2 \sum_{\text{1,3 distances}} F_{ij} (q_{ij} - q_0)^2 + \sum_{\text{1,3 distances}} F'_{ij} q_{ij} (q_{ij} - q_0) \\
 & + 1/2 \sum_{\substack{\text{tetrahedral} \\ \text{angles}}} \frac{3\kappa}{\sqrt{8}} (\theta_{ijk} - \theta_0)^2 + \sum_{\substack{\text{tetrahedral} \\ \text{angles}}} \frac{\kappa}{\sqrt{2}} (\theta_{ijk} - \theta_0)(\theta_{ijl} - \theta_0) + \sum_{\text{H-bonds}} E_{hb}
 \end{aligned}$$

where r_{ij} = the bond length between atoms i and j , θ_{ijk} = the bond angle between atoms i, j, k , q_{ij} the 1,3 distance between nonbonded atoms i and j , $\Delta\pi_{ijkl}$ = the out-of-plane bending (opb) coordinate defined by $\Delta\gamma \sin \theta_{jik}$ (γ being the angle between the bond $i-l$ and the plane jik), $\Delta\tau_{ij}$ the internal rotation coordinate around the bond $i-j$, K_b the bond-stretching force constant, H_{θ} , F and F' the attractive bond angle force constant, the repulsive bond angle force constant and the geminal force constant, respectively, κ the Kappa (internal tension) force constant, γ and Y force constants associated with the opb coordinate and the torsional coordinate, r_0 , θ_0 and q_0 are equilibrium parameters.

Inspection of this equation reveals that it is a rather complicated potential energy function. Slight differences with CHARMM and AMBER appear in the definition of the out-of-plane bending potential (sinus term)(ref.4), the dihedral torsional potential (harmonic approximation) and Van der Waals interactions (we use a Buckingham potential instead of a Lennard Jones potential). More striking divergences occur , however , for bond angles and hydrogen bonds. The term representating bond angles contains the attractive expression but also changes in the distance between the first and the third atom of an angle., the latter contribution being neglected in CHARMM and AMBER. This definition is by no means subtle. It should be emphasized that vibration frequencies of polyatomic molecule cannot be calculated satisfactorily without taking account for the atomic repulsion. The F' term ($F' = -0.1 F$ for a r^{-9} repulsion) does not have any effect

on vibration frequencies but is required in order that all the internal coordinate displacements, i.e. the q_i 's are zero and the energy is a minimum for $q_i = 0$, $i=1,2,\dots,3N-6$. But this argument is only valid if the coordinates are independent. If redundant coordinates are included (definition of six bond angle coordinates for the Methane molecule whereas there exists only five corresponding vibration frequencies), the argument must be revised by using the explicit redundancy condition given to the second order, and Kappa terms must be added. It should be emphasized that neither CHARMM nor AMBER takes into account the redundancy condition. The Kappa terms defined in the empirical potential energy function correspond to the case of a perfect tetrahedral.

The hydrogen bonding potential energy function (ref.5) is a combination of Van der Waals interactions and an explicit harmonic function, composed of terms representing bond length, bond angles and torsions, viz for a peptide hydrogen-bonding:

$$V_{hb} = \sum_{\substack{\text{nonbonded} \\ \text{pairs } r < 6 \text{ \AA}}} A \exp(-Br) - C/r^6 + 1/2 K_{O..H} (\Delta r(O..H))^2 + 1/2 H_{\alpha} r(C=O) r(H..O) (\Delta \alpha)^2 \\ + 1/2 H_{\beta} r(N-H) r(H..O) (\Delta \beta)^2 + 1/2 Y (\Delta \tau)^2 + 1/2 Y' (\Delta \tau')^2 + 1/2 Y'' (\Delta \tau'')^2$$

where α and β are the C=O..H and N-H..O bond angles, respectively, $\Delta \tau$, $\Delta \tau'$ and $\Delta \tau''$ the internal rotation displacements about the C=O, O..H and N-H bonds respectively. The energy is essentially dominated by the bond-stretching term. The parametrization of CHARMM and AMBER is different. The hydrogen bond energy is dominated by its electrostatic interactions. This representation yields potential surfaces in overall good agreement with scaled *ab initio* surfaces but is confronted to the reproduction of hydrogen-bonding vibrations in the liquid state. Therefore we do think that the best potential energy function should include an explicit bond stretching term, Van der Waals and electrostatic interactions.

The use of the harmonic potential energy function in the equations of motion leads to solutions that give harmonic molecular vibrations. The corresponding frequencies may be compared with experimental data obtained in the Infrared, Raman and Resonance Raman spectra. The force constants or parameters used in molecular mechanics studies are commonly optimized to get the best fit of calculated and experimental properties such as geometries, conformational energies and heats of formation. Very recently, it was rediscovered that vibrational spectroscopy can be used with great accuracy to derive force field parameters (refs.6,7). Parameters are most closely associated with the vibrational frequencies and their isotope effects. Moreover Coriolis coupling constants, centrifugal distortion constants, mean amplitudes of vibration and rotation-vibration coupling constants also give information on force field parameters. When we are dealing for example with the Cl, Br and I heavy atom vibrations, the Urey-Bradley-Shimanouchi force field is a good approximation. For hydrogen atom vibrations, this force field often gives frequencies which are far from the observed ones and for example a stretching-stretching interaction

constant has to be introduced in order to explain the high frequency spectra of methyl or methylene groups. In the same way a bending-bending interaction constant is needed in order to assign the mid frequency spectral range. Introducing such parameters leads to the so-called Modified Urey-Bradley-Shimanouchi force field (MUBFF). MUBFF is generally close to the General force field. For CH₃-X group the MUBFF consists of nine force constants : K(CH) , K(CX),H(HCH), H(HCX),F(HCH),F(HCX), internal tension, the stretching-stretching constant, and the bending-bending constant which are obtained from nine experimental frequencies. But the nine normal vibrations of CH₃-X are usually called the CH₃ totally symmetric stretching (TS) , the CH₃ symmetric deformation (SD), the CH₃ degenerate deformation (DD,DD') , the CH₃ degenerate rocking (DR,DR') and the CX stretching vibrations.

The suitable choice of basis coordinates in the expansion of V should correspond to these modes. The force constants on the basis of symmetry coordinates (local symmetry force field LSFF or group coordinate force field GCFF) are different from each other as shown on Table 1 where the MUBFF , the LSFF and their relationship together with the definition of coordinates are given.

TABLE 1
Force constants in the local symmetry force field.
Starting point : Urey-Bradley-Shimanouchi force field

stretchings(mdyne/A)

$$K(CC) = 2.563 \quad K(CH) = 4.301 \quad \text{CH}_3 \text{ group}$$

$$K(CH) = 3.936 \quad \text{CH}_2 \text{ group}$$

attractive and repulsive bond-angle bendings. (mdyne/A)

$$\begin{array}{ll} H(CCC) = 0.287 & F(CCC) = 0.369 \\ H(HCH) = 0.332 & F(HCH) = 0.279 \quad \text{CH}_2 \text{ group} \\ H(HCC) = 0.191 & F(HCC) = 0.537 \quad \text{CH}_2 \text{ group} \\ H(HCH) = 0.378 & F(HCH) = 0.195 \quad \text{CH}_3 \text{ group} \\ H(HCC) = 0.208 & F(HCC) = 0.385 \quad \text{CH}_3 \text{ group} \end{array}$$

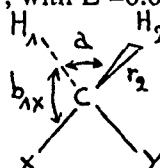
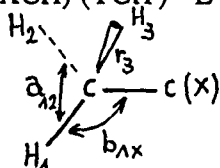
intramolecular tensions

$$\kappa(\text{CH}_3) = 0.025 \quad k(\text{CH}_2) = 0.054$$

interaction between CH₂ bendings (mdyne.A)

X-CH₂-Y group

$$2V = L (XCH) (YCH') - L (XCH) (XCH') , \text{ with } L = 0.011 \text{ mdyne.A}$$



5 atoms, 6 bending coordinates, 4 stretching coordinates ==> 9 independant coordinates

Definition of the symmetrical coordinates (ip means in plane and op out of plane)

$$\text{Redundancy: } S = 1/\sqrt{6} (a_{12} + a_{13} + a_{23} + b_{1x} + b_{2x} + b_{3x})$$

$$TS = 1/\sqrt{3} (r_1 + r_2 + r_3) \quad \text{ip}$$

$$DS = 1/\sqrt{6} (2r_1 - r_2 - r_3) \quad \text{ip}$$

$$DS' = 1/\sqrt{2} (r_2 - r_3) \quad \text{op}$$

$$\begin{aligned}
 SD &= 1/\sqrt{6} (a_{12} + a_{13} + a_{23} - b_{1x} - b_{2x} - b_{3x}) & \text{ip} \\
 DD &= 1/\sqrt{6} (2a_{12} - a_{13} - a_{23}) & \text{ip} \quad DD' = 1/\sqrt{2} (a_{12} - a_{13}) & \text{op} \\
 DR &= 1/\sqrt{6} (2b_{1x} - b_{2x} - b_{3x}) & \text{ip} \quad DR' = 1/\sqrt{2} (b_{2x} - b_{3x}) & \text{op} \\
 &\text{CX stretching}
 \end{aligned}$$

$$\begin{aligned}
 \text{Redundancy: } S &= 1/\sqrt{6} (a + b_{1x} + b_{2y} + b_{2x} + b_{1y} + XCY) \\
 SS &= 1/\sqrt{2} (r_1 + r_2) & AS &= 1/\sqrt{2} (r_1 - r_2) \\
 SC &= 1/\sqrt{20} (4a - b_{1x} - b_{1y} - b_{2x} - b_{2y}) & \text{ip} \\
 WA &= 1/\sqrt{6} (b_{1x} + b_{2x} - b_{1y} - b_{2y}) & \text{ip} \\
 TW &= 1/\sqrt{2} (b_{1x} - b_{2x} - b_{1y} + b_{2y}) & \text{op} \\
 RO &= 1/\sqrt{2} (b_{1x} - b_{2x} + b_{1y} - b_{2y}) & \text{op} \\
 &\text{CX stretching} \quad \text{CY stretching} \quad \text{XCY bending}
 \end{aligned}$$

the local symmetry force field
For a Methyl group

$$\begin{aligned}
 F(SD) &= 0.550 H(\text{HCH}) + 0.238 F(\text{HCH}) + 0.839 H(\text{CCH}) \\
 &\quad + 0.322 F(\text{CCH}) + 0.530 \kappa(\text{CH}_3) = 0.584 \text{ mdyne. A} \\
 F(DD) &= 1.188 H(\text{HCH}) + 0.475 F(\text{HCH}) - 0.177 \kappa(\text{CH}_3) = 0.534 \text{ mdyne.A} \\
 F(DR) &= 1.679 H(\text{CCH}) + 0.645 F(\text{CCH}) - 0.177 \kappa(\text{CH}_3) = 0.593 \text{ mdyne.A}
 \end{aligned}$$

For a Methylene group

$$\begin{aligned}
 F(SC) &= 0.950 H(\text{HCH}) + 0.380 F(\text{HCH}) + 0.336 H(\text{CCH}) \\
 &\quad + 0.129 F(\text{CCH}) - 0.318 \kappa(\text{CH}_2) = 0.538 \text{ mdyne. A} \\
 F(WA) &= 1.679 H(\text{CCH}) + 0.645 F(\text{CCH}) + 0.530 \kappa(\text{CH}_2) - 2L = 0.674 \text{ mdyne.A} \\
 F(RO) &= 1.679 H(\text{CCH}) + 0.645 F(\text{CCH}) + 0.530 \kappa(\text{CH}_2) + 2L = 0.717 \text{ mdyne.A} \\
 F(TW) &= 1.679 H(\text{CCH}) + 0.645 F(\text{CCH}) + 0.884 \kappa(\text{CH}_2) = 0.619 \text{ mdyne.A} \\
 F(SS) &= 4.735 \text{ mdyne/A} \quad F(AS) = 4.576 \text{ mdyne/A}
 \end{aligned}$$

The force constants of functional groups of organic molecules are usually influenced by the nearest neighbouring groups. They may be transferred to molecules having the same groups. For ring molecules, the redundancy conditions become complex. However it is always possible to define local symmetry coordinates of rings. A nice example is given by the benzene molecule for which UBSFF has shown strong limitations. Table 2 gives LSFF parameters as compared with published data.

TABLE 2
In plane vibrations of benzene

force constant (mdyne/A)		calculated frequencies (cm-1)	this work	MP2/6-31G* (ref.8)
A _{1g}				
	F11	993	7.6385	7.638
	F12		0.	0.135
	F22	3063.6	5.0912	5.625

A2g	F11	1366.4	0.5151	0.903
B1u	F11	1010.	0.8571	0.639
	F12		0.	-0.163
	F22	3057.1	5.0674	5.562
B2u	F11	1309.4	4.0673	4.049
	F12		-0.25	0.340
	F22	1146.	0.4710	0.864
E2g	F11	1178.3	0.54	0.838
	F12		0.	-0.106
	F13		-0.45	0.292
	F14		-0.153	-0.139
	F22	3055.7	5.077	5.571
	F23		0.	0.076
	F24		0.	0.029
	F33	1599.5	6.6546	6.938
	F34		-0.63	-0.449
	F44	606.9	0.933	0.942
E1u	F11	1035.	5.356	7.568
	F12		-0.18	0.249
	F13		0.	0.176
	F22	1479.5	0.5572	0.973
	F23		0.	0.008
	F33	3064	5.1052	5.60

Out of plane vibrations of benzene

Force constants and frequencies

B2g	F11	996.9	0.3845
	F12		-0.086
	F22	701.2	0.3595
E2u	F11	969.6	0.355
	F12		-0.095
	F22	401.7	0.367
E1g	F11	843.4	0.5234
A2u	F11	667.1	0.382

DERIVING EMPIRICAL POTENTIAL FUNCTION PARAMETERS FROM VIBRATIONAL SPECTROSCOPY.

There are several difficulties in fitting the vibrational spectra. If n_{exp} vibrational frequencies have been obtained, we are able to calculate n_{cal} normal modes frequencies. In principle it can be shown that there are 2^n equally good force fields giving 2^n sets of calculated frequencies with completely different assignments. Furthermore if N is the number of atoms in the molecule, the number of force constants which may be varied to improve the calculated frequencies increases as $(3N-6)(3N-5)/2$ while the number of observed frequencies is only $3N-6$. There are several ways to overcome these difficulties. They will be described in the following paragraphs where results obtained for different classes of molecules with biological importance will be presented.

A vibrational force field for peptides and proteins , and the use of the transferability of force constants, isotopic substitutions , mean-square displacements of atomic distances and crystallographic temperature factors as reliability criteria.(ref.9)

In order to setup parameters of the potential energy function for proteins , we have performed normal modes calculations of a series of small molecules in the crystalline state. The following molecules , i.e. Urea (ref.10), N-Methyl-Acetamide (NMA)(ref.11), N-Acetyl-L-X-N-Methylamide (X= Ala , Phe)(ref.12) have been studied whereas the molecules with X= Pro, Trp, Tyr and His are now in progress.

It is generally assumed that the principle of the transferability of force constants for a same series of molecules is applicable. Table 3 gives the calculated Amide vibration frequencies of the N-Acetyl-L-Alanine-Methylamide molecule from the molecular force field of the NMA molecule.

TABLE 3

obs. RAMAN freq.(cm-1)	obs. I.R. freq.	calculated freq.	Assignments
1667	1670	1656	AMIDE I C=O st.
1653		1639	
	1620	1565	AMIDE II N-H ipb.
	1570	1548	
1325	1320	1333	AMIDE III C-N st, C-C st
1300		1303	
	855	850	AMIDE V N-H opb
	768	743	
696	695	706	AMIDE IV C=O ipb
640		628	
593		617	AMIDE VI C=O opb
522	520	572	

where st= stretching , ipb= in-plane bending , opb = out-of-plane bending

Inspection of this table reveals that the principle is valid for most vibrations, exception of the C=O out-of-plane bending vibration, C=O group being near the side-chain. Accordingly specific interactions between this group and the side-chain are required in order to get better agreement. For all molecules the standart deviation of the order of 5cm-1 between the observed and computed frequencies is achieved. In addition to the frequency criterion, additional criteria have been used.

First in order to ensure the adequateness of the potential energy distribution for each vibration, the effects of isotopic substitutions are examined. Table 4 compares some observed characteristic frequencies of urea d-4 molecule with the calculated vibrations.

TABLE 4

Obs. freq.	calc. freq	Assignments.
2597	2583	ND2 asymmetric stretching
2438	2419	ND2 symmetric stretching
888	887	ND2 rocking
375	375	N-D out-of-plane bending

Then the mean-square displacements of atomic distances can be used. Table 5 points out the good agreement between the observed and the theoretical value for the N-H bond stretching coordinate displacement.

TABLE 5

Electron diffraction value	calculated value
0.0053	0.0051 (A ²)

The temperature factors can be also of great help when low frequency range spectra do not reveal all the lattice vibrations and all the internal vibrations associated with torsional coordinates. Inspection of the Table 6 shows a rather good agreement between observed and computed Debye-Waller factors for some atoms of N-Acetyl-L-Alanine-Methylamide. C₁H₃CO₁N₁HCH(CH₃)CONHCH₃.

TABLE 6

Atom-type	Observed values	calculated values (A ²)
C ₁	4.6 (1.0)	6.7
C(=O)	7.2 (1.2)	5.4
N ₁	4.8 (0.8)	5.2
O ₁	6.5 (0.8)	7.7

Finally it should be emphasized that the calculation of IR and neutronic vibrational intensities are in progress. The fundamental conclusion of these different studies is that the set of parameters can be quite different from those

utilized by CHARMM or AMBER. Table 7 gives an order of idea of the difference between some bond-stretchings force constants of NMA and the energy barriers to internal rotations about the bonds of NMA.

TABLE 7

our results		CHARMM	AMBER
		Kb (Kcal/mol A ²)	
C-C	155	187	335
N-CT	316	261	355
		E (Kcal/mol)	
N-C	17.3	20.	20.
C-C	1.1	0.	0.13
N-CT	0.18	0.	0.

So as to assess the importance of different potential energy functions on the calculation of dynamic properties of proteins, a harmonic dynamics simulation of BPTI (Bovine Pancreatic Trypsin Inhibitor) was carried out by using the Urey-Bradley-Shimanouchi force field.(ref.9).The density of vibrational states in the region 0 to 200 cm⁻¹ and the atomic fluctuations are different from the results obtained by previous harmonic dynamics simulations using a different empirical potential energy function. There can exist an important deviation for the lowest frequency mode, as it can be seen Table 8.

TABLE 8

Authors	Number of degrees of freedom	frequency (cm ⁻¹)
this work	1040	15.2
Brooks (ref.13)	1740	3.1
Cusack et al.(ref.14)	1740	11.0
Cusack et al (ref.14)	2712	9.6

Examination of the rms fluctuations of the Lys-15 residue reveals also a different dynamical behaviour(Table 9). This residue is of particular importance because there is a conformational transition around the C β -C γ bond with the fixation of the Inhibitor to the enzyme molecule. Accordingly the understanding of the mechanism of molecular recognition between two molecular entities depends upon the potential energy function

TABLE 9

Authors	Atom-type averaged rms fluctuations of Lys-15 residue.(A)	
	Type	
	Main-chain	Side-chain
this work	0.43	0.5
Brooks (ref.13))	0.7	1.5

Validation of force field using UV Resonance Raman Intensities. Application to Nucleic Acid Bases.

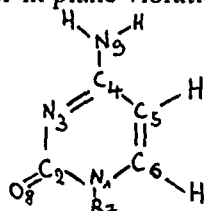
When absorption occurs during a $\Pi \rightarrow \Pi^*$ excitation, changes in the electronic distributions lead to a modification of the equilibrium position of the atoms in the excited state. The equilibrium position of atoms in the excited state can then be expressed by a vector $X_e \{X_1, \dots, X_{3n}\}$ whose elements are functions of the ground state cartesian displacement coordinates and X_e represents the displacement of the equilibrium atomic coordinates of the electronic excited state from the equilibrium atomic coordinates of the ground state..

In another way we define R_e as the set of internal valence bond displacement coordinates between the ground and the excited state $\{R_{e1}, \dots, R_{e3n-6}\}$, $\Delta_e \{\Delta_{e1}, \dots, \Delta_{e3n-6}\}$ is the excited state displacements of the normal coordinates and is related to R_e by $R_e = L \cdot \Delta_e$, L being the matrix of the ground state vibrational eigenvectors. $L = \{L_{i1}, \dots, L_{ij}\}$ is the participation of each internal coordinate R_{i1}, \dots, R_k to the normal mode. The inverse transformation leads to $\Delta_e = L^{-1} \cdot R_e = L^{-1} \cdot B \cdot X_e$. Using *ab initio* quantum mechanical or semi-empirical calculations, it is possible to calculate the cartesian coordinates of the minimum of potential energy of the excited and ground electronic states. From normal mode calculations, we can obtain the L , and L^{-1} matrices. However, these minimizations are time computer expensive. In the case where the displacement of the minima of potential energy between the ground state and the excited state is weak (rigid molecule and first electronic transitions), there is a linear relationship between the change of bond order of C-C and C-N bonds and the change in their bond length (ref.15). Then we have $\Delta_e = L^{-1} \cdot R_e = L^{-1} \cdot (\Delta b.o)_e$ where $\Delta b.o$ means the change in bond order, and R_e the change in bond length from the ground to the excited state. The intensity I_i of the totally symmetric Raman band depends of the displacement of the normal mode and is related to $I_i = k \omega_i^2 (\Delta_{ei})^2$, ω_i vibrational frequency of the i th normal mode, k is a constant, Δ_{ei} displacement of the i th normal mode in the excited state. It is then possible to test the molecular force field by calculating in the same way the $3n-6$ normal modes frequencies and the $3n-6$ corresponding intensities. Setting the intensity of the most intense band to 10, it is possible to calculate the relative intensities of the others bands.(refs.15,16).

Figure 1 displays experimental and calculated (vertical lines) Resonance Raman Intensities of 5'CMP at 266 and 213 nm. Table 10 gives the corresponding in-plane general force field.

TABLE 10

General Valence Force Field for in plane vibrations of 1-Methyl Cytosine.



Force constants are in mdyne/A for stretchings and in mdyne.A for bendings

v stretching, δ ring in plane bending., vsk ring skeletal stretching

N1-C2	8.86	(8,2,3)	1.023	$\delta(\text{CNH}_2)$	0.286
C2-N3	6.30	(2,3,4)	1.189	$\delta(\text{NH}_2)$	0.425
N3=C4	5.34	(3,4,5)	0.62	$\nu\text{C}=\text{O}$, vsk	1.637
C4-C5	5.042	(3,4,9)	1.412	$\nu\text{C}=\text{O}$, $\delta(\text{C}=\text{O})$	1.363
C5=C6	9.56	(9,4,5)	1.412	vsk, $\delta(\text{C}=\text{O})$	0.304
C6-N1	7.34	(4,5,6)	0.62	v, v meta	-0.204
N1-H	5.604	(4,5,10)	0.37	v, v gem	0.851
C2=O8	10.55	(6,5,10)	0.37	v, v para	0.88
C4-N9	6.57	(5,6,1)	0.62	δ , vsk	0.60
C-H	5.28	(5,6,11)	0.412	vsk, $\delta(\text{N-R})$	0.193
N9-R	6.00	(1,6,11)	0.412	vsk, $\delta(\text{C-H})$	0.445
(6,1,2)	1.844			$\nu(\text{N-C})$, $\delta(\text{C-H})$	0.058
(6,1,7)	0.482			vsk, $\delta(\text{CNH}_2)$	0.45
(2,1,7)	0.482			$\delta(\text{C-H})$, $\delta(\text{C-H})$	-0.036
(1,2,3)	1.62			$\delta(\text{C-H})$, $\delta(\text{N-H})$	0.1
(1,2,8)	1.534			$\delta(\text{C}=\text{O})$, $\delta(\text{N-H})$	0.0

Using the calculated force field from intensities of the UVRR for the guanosine and cytosine moieties, and using force constants derived for ribose from Baret et al. (ref.17) work's onto poly (dA) poly (dT) and poly d(AT) poly d(AT) in the B form, we have in a first approximation calculated the normal modes for the Z form of poly (dGdC) poly(dGdC) as was first observed in X rays by Wang et al. (ref.18) and observed in Raman spectroscopy by Thamann et al (ref.19) in the solid state. The characteristic band of Z form of DNA, observed at 617 cm^{-1} was calculated at 625 cm^{-1} as arising from a normal mode coupling between the phosphate ribose moiety and the glycosidic-guanosine linkage(ref.20).

A second example of application is the transfert of force constants to the calculation of normal modes of the crystal of the base pair model 1 methyl cytosine-9 ethyl guanine (ref.21). The purpose of this work was to study the hydrogen bond effects on normal modes. Particular interesting is the potential energy distribution of the calculated 780 cm^{-1} normal mode of the cytosine group mainly CH group out of plane vibrations and the small participation of out of plane C=O group, which was observed in UV Resonance Raman spectrometric experiments at 295 nm by Chinsky et al. (ref.22).

Determination of parameters for carbohydrates and oligosaccharides.

First a normal coordinate analysis is performed for both α - and β -D-glucose molecules in the crystalline state. The intra- and inter-molecular potential energy function used is that described previously. Above 200 cm^{-1} the potential energy distribution of each vibration determined by previous calculations for an isolated molecule (refs.23-27) is refound. Comparison with Hineno's works(ref.26) on β -D-glucose, (calculation of an isolated molecule with Urey Bradley Shimanouchi Force Field) shows important improvements.

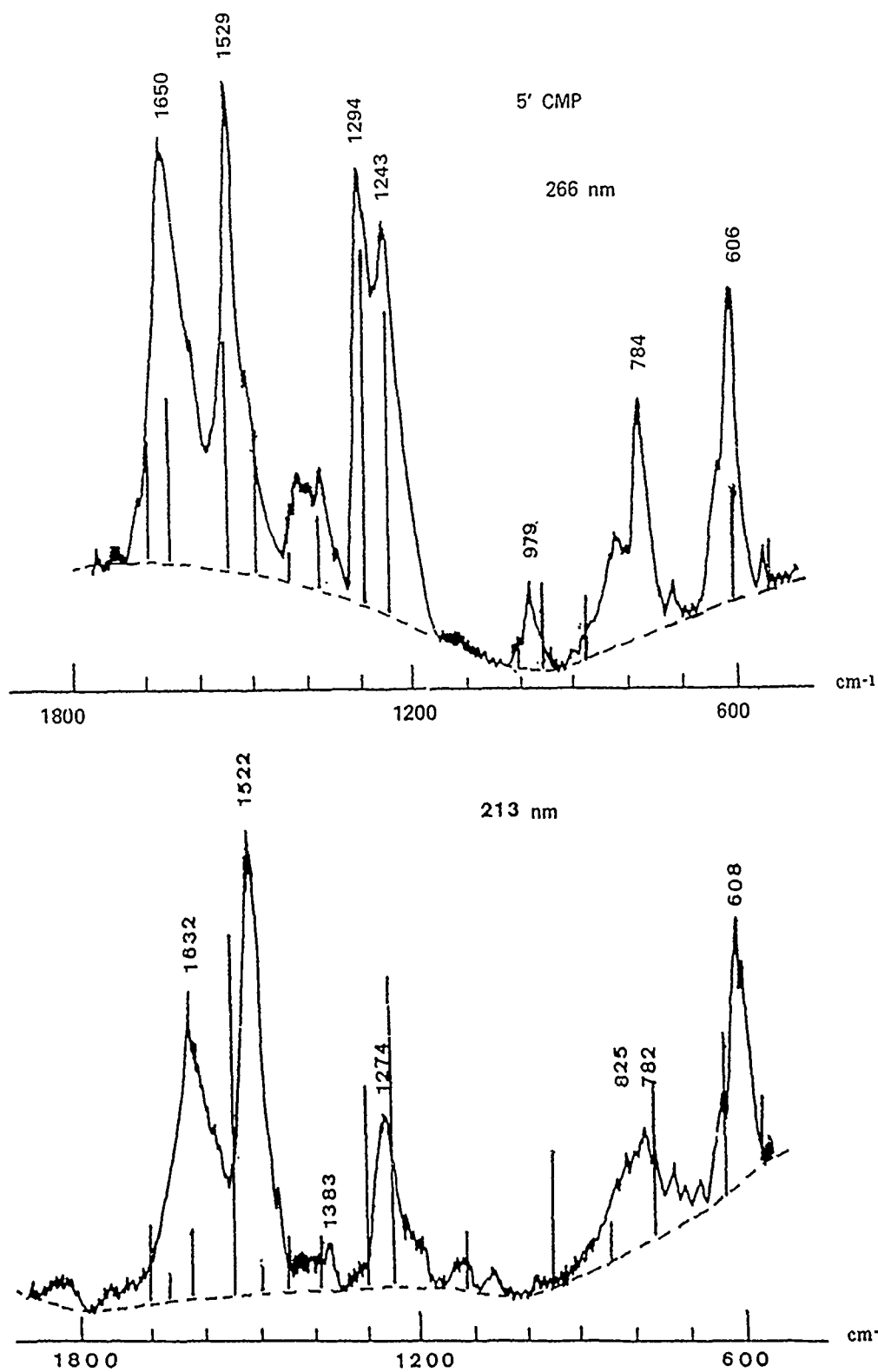


fig.1 Resonance Raman spectra of 5' CMP.

The anomeric characteristic groups (C_1HOH , C_6H_2OH) were treated separately from the ring moiety. In a counterpart, we have distinguished the force constants for the ring, the C_6H_2OH group and the anomeric carbon group. We have found that the values of the force constants decrease considerably for the atomic groups involved in the network of hydrogen-bondings; as shown in Table 11.

TABLE 11

this work	ref.26 units are in mdyne/A
H(COH) = 0.256	0.413
F(COH) = 0.586	0.750
H(C_6OH) = 0.340	
H(C_6OH) = 0.690	
H(HCO) = 0.250	0.279
F(HCO) = 0.710	0.850
H(HC_1O) = 0.250	
F(HC_1O) = 0.630	
H(HC_6O) = 0.190	
F(HC_6O) = 0.540	

For the others force constants associated with the bond angle deformations, no major changements were observed. The force constants associates with the torsional coordinates are slightly different because of the treatment of hydrogen-bondings (see Table 12).

TABLE 12

this work	ref.26 units in mdyne.A
YCCring = 0.150	0.090
YCCside = 0.125	0.090
YCOring = 0.0615	0.100
YCOside = 0.065	0.350

It is important to point out that approximately the same set of parameters for the α -D-glucose has been applied (differences existing in the geometry modifications) in order to explain characteristic frequency shiftings between the two anomeric conformers. In the spectral range below 200 cm^{-1} , the usefulness of electrostatic contributions, charge distribution computed by AM1 quantum mechanical procedure (ref.28) with a dielectric constant of 3, and the reliability of the treatment of hydrogen-bondings by an explicit function is clearly demonstrated by the perfect agreement between computed and observed vibrations for both anomers. Then this set of parameters has been used for others monosaccharides, i.e. both anomers of galactose, methyl- α -D-glucose and N-acetyl-glucosamine. The assignment of all the observed bands is possible, and it seems that this set of parameters is reliable and transferable. In a near future, molecules as methyl- β -D-glucose, methyl- α -D-galactose, methyl- β -D-galactose, methyl- α -D-mannose, α -L-fucose will be studied in order to explain the exo-anomeric effect and to assign the frequencies of characteristic groups (in addition of those obtained for both glucose molecules and deuterated analogs).

To complete the set of parameters the study of different disaccharides, thus different types of glycosidic linkages, will be investigated. The disaccharides as maltose and cellobiose will give parameters for the kind of linkage (α or β). whereas the disaccharides as trehalose, sophorose, laminarabiose and gentiobiose will provide parameters for each glycosidic linkage position. The total set of parameters will be used to perform dynamics simulation of polysaccharides.

CONCLUSION

The vibrational spectroscopy (frequencies and intensities of vibrational bands) is a very powerful technique to derive force fields parameters. Precise internal rotation barriers values are also obtained. Modifications of the commonly used potential energy function have to be taken into account according to normal modes calculation of model compounds with biological interest in the crystalline state. The obtained parameters, together with the function, may be introduced in the various macromolecular mechanics programs dedicated to biomolecular structure and dynamics determination. In the near future, much attention has to be focused on molecules with therapeutical interest (rings, π electrons systems) in order to develop accurate molecular force fields on the basis of local symmetry (group) coordinates. In this case the determination of parameters has to be done for each structural class of compound using both the strategy defined for amino-acids and nucleic acid bases and quantum mechanical results.

REFERENCES

- 1 T. Shimanouchi, in "Physical Chemistry" volume 4, "Molecular properties", chapter 6, 233-306, Academic Press New York (1970)
- 2 B.R. Brooks, R. Brucolleri, B. Olafson, D. States, S. Swaminathan, M. Karplus, *Comput. Chem.*, 4, (1983), 187.
- 3 P.K. Weiner, P.A. Kollman, *J. Comput. Chem.*, 2, (1981), 287.
- 4 T. Miyazawa, K.S. Pitzer, *J. Chem. Phys.*, 30(4), (1959), 1076.
- 5 M. Suzuki, T. Shimanouchi, *J. Mol. Spect.*, 28, (1968), 394.
- 6 L. Nilsson, M. Karplus, *J. Comput. Chem.*, 7(5), (1986), 591.
- 7 S.J. Weiner, P.A. Kollman, D.T. Nguyen, D.A. Case, *J. Comput. Chem.*, 7(2), (1986), 230.
- 8 H. Guo, M. Karplus, *J. Chem. Phys.*, 89(7), (1988), 4235.
- 9 P. Derreumaux, PhD Thesis, (1988), University of Lille, France.
- 10 P. Derreumaux, P. Lagant, G. Vergoten, submitted to *J. Comp. Chem.*
- 11 P. Derreumaux, G. Vergoten, G. Hong, M. Karplus, to be submitted to *J. Comp Chem.*
- 12 P. Derreumaux, G. Vergoten, M. Karplus, to be submitted to *J. Comp Chem.*
- 13 B.R. Brooks, M. Karplus, *Proc. Nat. Acad. Sci. U.S.A.*, 80, (1983), 6575.
- 14 S. Cusack, J. Smith, S. Finney, B. Tidor, M. Karplus, *J. Mol. Biol.*, 202, (1988), 903.
- 15 W.L. Peticolas, D.P. Strommen, V. Lakshminarayanan, *J. Chem. Phys.*, 73(9), (1980), 4185.

- 16 W.L. Peticolas, P.Lagant, G. Vergoten, Meeting of the SPIE, Los Angeles, USA, January 1986.
- 17 J.F. Baret, G.P. Carbone, J. Sturm, J. Raman Spect., 8(6), (1979), 291.
- 18 A.J.H. Wang, G.J. Quigley, F.J. Kolpak, J.H. Van der Marel, J.H. Van Boom, A. Rich, Science, 211, (1981), 171.
- 19 T.J. Thamann, A.J.H. Lord, H.J. Wang, A. Rich, Nucleic Acids Res., 9, (1981), 5443.
- 20 G. Vergoten, P. Lagant, W.L. Peticolas, Y. Moschetto, I. Morize, M.C. Vaney, J.P. Mornon, J. Molecular Graphics, 4(4), (1986), 187.
- 21 P. Lagant, G. Vergoten, P. Derreumaux, R. Dhennin, submitted to J. Raman Spect.
- 22 L. Chinsky, B. Jolles, A. Laigle, P.Y. Turpin, Biopolymers, 23, (1984), 1931.
- 23 P. Vasko, Carbohydrate Res., 23, (1972), 407.
- 24 J. Cael, Carbohydrate Res., 32, (1974), 79.
- 25 V. Andrianov, Zh. Strukt. Khim., 21(1), (1980), 35.
- 26 M. Hinenno, Carbohydrate Res. 56, (1977), 219.
- 27 J.P. Huvenne, G. Vergoten, P. Legrand, G. Fleury, J. Mol. Struct. 74, (1981), 169.
- 28 M. Dewar, J. Amer. Chem. Soc., 107, (1985), 3902.

DISCUSSION

SOUMPASIS - What are the differences between your and the Karplus group calculations of the normal modes of BPTI ?

VERGOTEN - The lowest frequency is higher than that calculated by Brooks et al (PNAS, USA 80 (1983), 6575) and Cusack et al (J. Mol. Biol., 202 (1988) 903). The atom type averaged rms fluctuations of the Lys.15 residue are smaller (0.43 Å for the main chain and 0.5 Å for the side chain instead of 0.7 and 1.5 respectively in Brook's paper which seem to be in better agreement with the strong stability of the trypsin-inhibitor complex.

STONE - Your calculations use a harmonic force field, so one should not expect close agreement with experiment, where anharmonic effects contribute. Can you comment on the effects of anharmonicity ?

VERGOTEN - In general the harmonic assumption is satisfactory enough in explaining most internal vibrations and reflecting the role of intermolecular hydrogen bonding on local vibrations. In some cases however observed frequencies cannot be reproduced by a "harmonic" calculation. In large scale motions, the cubic term in the Taylor expansion of the potential energy has to be included. This is however done in the treatment of small molecules.

DYMEK - Which aspects of your approach do you consider to be contributing most to the fact that you do not obtain negative eigenvalues in the vibrational analysis ?

VERGOTEN - The origin of negative eigenvalues in normal modes analysis may be double : either the starting geometry is very bad or the potential energy (and especially torsional terms and hydrogen bond parameters) is physically meaningless. For systems for which X ray data are available the second argument is the one to be taken into consideration. Usually when negative eigenvalues are found, people continue to minimize their structure until the problem disappears. It seems to me (and that is what we are doing), that in this case it is more reasonable to try to fit the X ray structure by modifying the potential energy terms.

SMITH - With reference to the previous comment on negative eigenvalues obtained by the Karplus group at Harvard - These were seen for the first analysis, of the bovine pancreatic trypsin inhibitor (Brooks, B. and Karplus, M. P.N.A.S. (U.S.A) 80 6571 (1983)). Due to increased computer power, this is no longer a problem and protein calculations of this type reach energy minima, without negative eigenvalues. Some of these analyses produce neutron spectra in good agreement with the experimental results, but the calculated low frequency density of states is very sensitive to long range electrostatic effects (Cusack et al, J. Mol. Biol. 202, 903 (1988), Smith et al, Physica B 156 & 157, 437 (1989)). Finally, one must remember that vibrational (as opposed to molecular mechanics) force fields are specialised and cannot give direct structural information, as the geometry used is assumed to be at an energy minimum.

AN INTRODUCTION TO NUMERICAL OPTIMIZATION

Claude LEMARECHAL

INRIA - BP 105 - 78153 Le Chesnay (France).

SUMMARY

We review the main ideas underlying numerical optimization, adopting the point of view of actually using an algorithm on a computer. We briefly describe some of the most useful methods, with emphasis on the unconstrained problems. Finally, we use an application in molecular biology for illustration.

I. GENERAL IDEAS

To solve an optimization problem is to find a set of variables, satisfying some given constraints, and minimizing a given cost-function. We consider only the so-called mathematical programming problem, in which the variables vary in \mathbb{R}^n , and there are finitely many constraints.

The solution (if there is one) is actually approximated, using an algorithm which constructs an iterative sequence of trial solutions. The user must provide a program which, for any given value of the variables, computes the corresponding values of the cost and/or constraints (and also their partial derivatives; we will return to this later). From this point of view, optimization is fairly similar to, say, nonlinear systems, or differential equations.

We explain the main ideas that are used to define optimization algorithms. They are best viewed in the unconstrained case, i.e. when the variables are allowed to take any value in \mathbb{R}^n . In this framework, the simplest problem is when the cost-function is quadratic, say

$$f(x) = \frac{1}{2} x^T A x + b^T x + c.$$

Here, A is the matrix of second derivatives of f , which actually does not depend on x . The problem then reduces to a linear system of equations: $Ax + b = 0$.

When f is not quadratic, its matrix of second derivatives varies with x . To be efficient, an optimization algorithm must somehow estimate this matrix, thus obtaining a quadratic function supposed to estimate f . This is the basis for Newton-like methods, conjugate gradient, etc... We demonstrate this approach and we explain how it can be adapted in various situations, such as when the problem is large-scale, or stiff, or when it originates from a least-squares problem, etc...

Finally, we show how the above ideas can be extended to constrained situations: first in the case of linear constraints (the feasible domain is a polyhedron), and also in the general case when the constraints are nonlinear. In the former case, the sequence of iterates is usually feasible, while it can hardly be so in the latter: it is only asymptotically that nonlinear constraints are eventually satisfied.

II. THE PROBLEM TO SOLVE

We will denote the problem of interest in this paper by

$$\begin{cases} \min f(x) & x \in \mathbb{R}^n \\ g_i(x) \leq 0 & i = 1, \dots, m \\ g_j(x) = 0 & j = m+1, \dots, m+p \end{cases} \quad (1)$$

In this notation, x represents the control variables; f is the cost-function; the g 's are the constraints – inequality constraints (the first m) and equality constraints (the last p) – restricting the set of feasible variables. An x satisfying all the constraints will be called feasible. Note that f is scalar-valued: there is only one objective function (this is very important; we do not consider multi-objective optimization). In other words, we want a value of $x \in \mathbb{R}^n$ satisfying all the constraints and making the objective function as small as possible.

Remark.

If, instead of \mathbb{R}^n , x were varying in a finite or discrete set, we would obtain a so-called problem of integer-programming, or of combinatorial optimization. If \mathbb{R}^n were actually an infinite-dimensional space, i.e. if x were a function, say of the time (instead of a familiar vector with finitely many coordinates), we would have an optimal control problem. None of these two cases will be studied here. ■

Depending on the type of constraints, several classes of problems must be considered. One has

1. Unconstrained problems, i.e. $m=p=0$. These problems are prototypical and serve as models to the others, with various types of constraints. As seen above, the case of a quadratic f is trivial: one then has to solve a linear system expressing that $\nabla f(x)=0$. When f is not that special, a solution can only be approximated, by an iterative algorithm.

2. Linearly constrained problems (the constraints are affine functions, say $g_i(x)=a_i^T x + b_i$, where $a_i \in \mathbb{R}^n$ and b_i is a number).

2.1. Problems with equality constraints only ($m=0$, $p>0$) which can actually be recast into class 1 (for example, use the constraints to extract $n-p$ independent, and unconstrained, variables).

2.2. Problems with inequality constraints ($m>0$, p arbitrary)

2.2.1. Linear programming, where f is affine (as well as the constraints); an intermediate case with combinatorial optimization, which requires a special

method: the simplex method or, since more recently, interior (Karmarkar) methods.

2.2.2. Linear-quadratic problems, where f is quadratic (and the constraints affine); hardly more difficult than 2.2.1, they can also be solved exactly.

2.2.3. All the others, which can only be solved iteratively.

3. General problems, where one distinguishes again

3.1. Problems with equalities only, of the same type as class 1 in that they have a comparable complexity.

3.2. Problems with inequalities, or general nonlinear programming.

It should be noted that, as soon as inequality constraints are added to a given problem, its theoretical complexity is drastically increased (there are potentially 2^m equality-constrained problems to solve: one does not know "a priori" which inequality will be active at the optimum). In practice, however, the situation is not that bad: it is usually not necessary to visit all these combinations.

In the above classification, several subcases must be distinguished depending on the form of the (nonlinear) functions involved.

a. Sum of squares: $f(x) = \frac{1}{2} \sum F_j(x)^2$, i.e. one wants to solve a system of equations $\vec{F}(x) = 0$ by the least-squares method. The vector \vec{F} need not be in \mathbb{R}^n but in some other space, usually larger (there are "too many equations"). Here again, a trivial case is when \vec{F} is affine (linear regression).

b. Nondifferentiable problems are those in which one of the functions, say f , has discontinuous first derivatives. Typical examples are when one chooses in a. above, instead of the ℓ_2 -norm:

- * either the ℓ_1 -norm, in which $f(x) = \sum |F_j(x)|$ (derivatives fail to exist when some $F_j = 0$)
- * or the ℓ_∞ -norm: $f(x) = \max |F_j(x)|$ (derivatives fail to exist when several F_j are maximal).

c. Large-scale problems: when n is large, essentially when it is impossible, or inconvenient, to compute or even store the n^2 second derivatives of f ; likewise, when it is impossible, or inconvenient, to store the Jacobian matrices of the constraints.

Finally, another subclassification is based on the available information concerning f and g . This is the subject of the next section.

III. GENERAL STRUCTURE OF AN OPTIMIZATION PROGRAM

When actually implemented on a computer, an optimization program is made of two very distinct parts. One is the algorithm properly said, which can be considered as the decision-maker: it builds the sequence $\{x_k\}$ of iterates, supposed to converge to an optimum. The second part defines the functions characterizing the problem: f and g ; it provides the decision-maker with the necessary informations allowing an intelligent construction of $\{x_k\}$. We will call the simulator this second part.

An essential feature of optimization problems is that the simulator gives purely point-wise information: the algorithm must content itself with numerical values of f , g , etc... at numerical values of x , and nothing more. In particular, no global information is available other than theoretical (continuity, differentiability...): the simulator should be thought of as a bunch of punched cards, say, involving a huge amount of coding, impossible to analyze (at least in the present state of computer science). This is common to other nonlinear problems such as, say, differential equations. A second feature (which is not shared by differential equations) is that the algorithm has to proceed by trials and errors: one is never sure in advance that a given x will be a good iterate.

A naive example will illustrate these two points: imagine a tennisman practicing his serve. He wants to optimize something like the final speed of the ball (when it reaches his adversary's area) and he can control initial conditions (velocity, direction, spin...). When these initial conditions are fixed, the ball follows a certain trajectory, to reach the final state one is interested in. Upon observation of this state, the player corrects his initial conditions, observes the resulting effect, and so on. To implement this system on a computer, one needs above all the differential equations that simulate the trajectory; then a tennis ground is no longer needed, provided that one has on hand an intelligent program to iterate over the initial conditions. One feels that important informations for this program will be, among others, the sensitivity of the final state with respect to the initial conditions (this sensitivity is more or less felt, at least qualitatively, by an experienced tennisman).

It is essential that the above decomposition "algorithm+simulator" be reflected in the general organization of the computer-program realizing an optimization: the algorithm is generally written by an applied mathematician (it is a full-time profession...); the simulator by a user from some other discipline: physics, chemistry, economics... and the two programmers may well never meet.

As suggested in the tennis-example above, the simulator can be more or less sophisticated and there are several cases:

- (i) At each x , the simulator merely computes $f(x)$ and the constraint-vector $g(x)$ (it is the least that can be asked from it!).
- (ii) It computes the partial derivatives as well, to obtain the gradient $\nabla f(x)$ and the Jacobian $\nabla g(x)$.
- (iii) In addition, it computes also the second derivatives, let $\nabla^2 f(x)$ ($\nabla^2 g$ is rarely useful).

Remark.

Note that in the case of linear functions, say $g_i(x) = a_i^T x + b_i$, no simulator is needed in principle, since the mere knowledge of a_i and b_i (a finite amount of data, instead of a subroutine) is sufficient to compute all the possible values of the corresponding function g_i . The case of quadratic functions is similar. Nevertheless, a simulator may sometimes be useful in this situation. For example, it may happen that gradient vectors such as a_i and/or Hessian matrices cannot be fully computed, or even stored; while a subroutine computing $a_i^T x + b_i$ might be easy to write and execute. ■

The most common case is (ii). We know of no situation in which the simulator returns more information than in (iii); yet, this latter case (iii) is rarely rewarding: much work is required from the simulator, which usually results only in a slight improvement of possible algorithms. As for (i) (no derivative), it should be avoided by all means: it only allows algorithms that are intolerably slow, if at all convergent — the most popular of which is the method of Nelder & Mead.

Remark.

Thus, a dialogue in which first derivatives are provided can be considered as the standard situation, realizing an appropriate balance between efficiency and complexity. It should be noted, however, that the requirement for the user to compute these derivatives is a serious drawback of numerical optimization. Although differentiation is always possible in theory, writing the corresponding code can be highly complex in practice. Examples are not rare where it requires several men.years. ■

IV. GENERAL STRUCTURE OF AN OPTIMIZATION ALGORITHM

In view of the limited information available from the simulator, an optimum can be detected via a trial and error scheme only (remember the tennisman: he is never sure that his next serve is going to be better than the present one). Thus, one has to use an algorithm, to generate a sequence of trial iterates, say $\{x_k\}$ (the serves of the tennisman), hopefully converging to an optimal solution. Such an optimization algorithm is usually based on the following ideas:

- A test-function $\varphi(x)$ is defined, which must be minimal at a solution of (I). For unconstrained problems, one takes of course $\varphi \equiv f$. For linear constraints, it is usual to organize the calculations so that the whole sequence $\{x_k\}$ is feasible (possibly after a Phase I aimed at finding an initial feasible iterate); then, one can take also $\varphi \equiv f$. Nonlinearly constrained problems are more delicate, since it is impossible in general to find a feasible x in a finite amount of calculations. In these cases, φ must be a compromise between f -values and violation of the constraints. Generally, one takes an exact penalty function

$$\varphi_r(x) = r f(x) + \sum_{i=1}^m \max[0, g_i(x)] + \sum_{j=m+1}^{m+p} |g_j(x)| \quad (2)$$

where r is a suitably small coefficient.

- At a given iterate x_k , a direction $d_k \in \mathbb{R}^n$ is computed. It is a descent direction, i.e. such that φ decreases locally along d_k : $\varphi(x_k + t d_k) < \varphi(x_k)$ for $t > 0$ small enough. To compute this direction, the differential information concerning (1) is useful, if not compulsory.
- Then a descent step is made. It consists of searching $t_k > 0$, to obtain $x_{k+1} = x_k + t_k d_k$ such that $\varphi(x_{k+1}) < \varphi(x_k)$. This search is by no means easy, because of the poor information provided by the simulator. It is here that the trial and error process must take place. This part of the algorithm is sometimes called the line-search.
- Furthermore, the stopping criterion is performed, to detect whether the current x_k is close enough to optimality. Keeping (2) in mind, this is based on checking the magnitude of the gradient of φ .

One iteration of a minimization algorithm at a given iterate x_k thus consists of performing two essential things: first compute the descent direction d_k , then compute the actual stepsize t_k . The problem of the direction is the more important one and will be considered in the next section. Here and now, we give a few indications concerning the line-search.

Given the starting point $x (=x_k)$ and the descent direction $d (=d_k)$, the simulator is called at a trial stepsize $x+td$. Depending on its answer, this t is accepted (and the next iterate is $x_{k+1} = x+td$), or corrected until it becomes convenient. Thus, a line-search is a trial-and-error process, defined by two things: what is a convenient stepsize, and how an inconvenient stepsize is corrected.

The stepsize is corrected on the basis of a safeguarded polynomial fit: in case (ii), one has on hand the value of φ and of $\nabla\varphi$ at each trial stepsize, including $t=0$ (the simulator has been called at the current iterate x_k). Most commonly, one uses these informations at the current t and at the previous t_- to define a cubic function of the single variable t , whose local minimum (when it exists) gives the next trial stepsize. In addition, one has on hand a "bracket", also made of two previous trials, inside which a convenient stepsize is known to lie. The cubic interpolator is then forced inside the bracket, which is thus contracted at each trial.

Defining the set of convenient stepsizes (i.e. the stopping criterion for the line-search) is by far the more important item: it must be tolerant enough to allow a fast stopping of the line-search, and strict enough to ensure convergence of (x_k) toward a minimum. Roughly speaking, the line-search is stopped if the current $t > 0$ is

- not too large, which is quantified by the fact that $\varphi(x+td) - \varphi(x)$ is sufficiently negative (if so is not the case, t becomes the right-end of the bracket)

— not too small, which is quantified by the fact that $\varphi'(x+td) - \varphi'(x)$ is sufficiently positive (the notation φ' means the derivative of φ , considered as a function of the single variable t ; when t is not too large and not too small, the line-search is finished; if t is not too large but too small, it becomes the left-end of the bracket).

Properly programmed line-searches thus accomplish an iteration in less than 2 or 3 trials.

V. USEFUL MINIMIZATION ALGORITHMS

Considering that all optimization algorithms use similar line-search-procedures, they differ only in the way the direction is computed. This direction is computed on the basis of theoretical considerations aimed at decreasing the function φ of (2). It is in contrast to the line-search, which works "a posteriori", on the firm ground of actual φ -values. The basic way of reasoning for the direction is as follows: one supposes that, close to x_k , φ has such or such form; then one asks what a suitable direction should be, under this simplifying assumption. Specifically, the objective function is assumed quadratic and the constraints (if any) linear; this places the problem in class 2.2.2., which can be solved exactly.

Remark.

We mention here that, in case of nonlinearly constrained problems, the relevant curvature is not that of f , but that of the Lagrange function: the curvature of the constraints must be taken into account as well.

When making this linear-quadratic assumption, the function of (2) becomes

$$\varphi_r(x) = r[f(x_k) + \nabla f(x_k) \cdot (x - x_k) + \frac{1}{2}(x - x_k)A(x - x_k)] + \sum \max[0, g_i(x_k) + \nabla g_i(x_k) \cdot (x - x_k)] + \sum |g_j(x_k) + \nabla g_j(x_k) \cdot (x - x_k)|$$

in which A represent the relevant curvature. Minimizing φ_r amounts to solving the problem with $n+m+p$ variables and $2(m+p)$ linear constraints

$$\begin{cases} \min r[\nabla f(x_k) \cdot (x - x_k) + \frac{1}{2}(x - x_k)A(x - x_k)] + \sum_{i=1}^m v_i + \sum_{j=m+1}^{m+p} v_j \\ v_i \geq 0 ; v_i \geq g_i(x_k) + \nabla g_i(x_k) \cdot (x - x_k) \\ v_j \geq g_j(x_k) + \nabla g_j(x_k) \cdot (x - x_k) ; v_j \geq -g_j(x_k) - \nabla g_j(x_k) \cdot (x - x_k) \end{cases} \quad (3)$$

whose solution can be considered as approximating a solution of (1) (if x_k is close to such a solution, and if r is small).

It is for efficiency that second order effects taken into account when computing the direction. As mentioned earlier, no definite increase of performances is usually to be expected from higher order considerations. On the other hand, a first order analysis is defi-

nately insufficient and gives drastically worse results. As a rule, an optimization algorithm is efficient if and only if it is based on a second order analysis.

Naturally, the second order information is not always available from the simulator, especially in the standard case (ii). Indeed, the whole issue when computing the direction is to identify the (necessary but unknown) second derivatives. We review below some of the most used methods, depending on the situation. For the sake of simplicity, we consider the unconstrained case only: the problem is to minimize the function f over the whole \mathbb{R}^n ; as mentioned earlier, there is no point in considering any special function φ , which is just f .

In case (iii), we can write

$$f(x_k + d) \simeq f(x_k) + \nabla f(x_k) d + \frac{1}{2} d \nabla^2 f(x_k) d$$

so it makes sense to take d minimizing this quadratic approximation, i.e. to solve

$$\nabla^2 f(x_k) d = -\nabla f(x_k). \quad (4)$$

We recognize the Newton method for solving $\nabla f(x) = 0$. The possibility of performing a line-search along the resulting direction is a great advantage over general systems of equations: stability is automatically obtained from the mere property $f(x_{k+1}) < f(x_k)$. Naturally, the Newton iterate is obtained with the stepsize $t=1$ (which would be the only stepsize "a priori" reasonable, if f were not known in addition to the vector-field ∇f); therefore $t=1$ can be used to initialize the line-search. All this is valid only if the matrix $\nabla^2 f$ is positive definite; otherwise, the Newton direction might not be downhill; furthermore, the Newton iteration would have a tendency to converge to a saddle-point of f , rather than a minimum. When $\nabla^2 f$ is positive definite, (4) is best solved by the Cholesky factorization of $\nabla^2 f$. We add here that, in the case of a large problem (when $\nabla^2 f$ is impractical), (4) can be solved by an iterative method (relaxation, conjugate gradient...) interrupted before convergence (so as to get a reasonable direction in a reasonable computing time). This gives truncated Newton methods.

Most usually, namely in case (ii), $\nabla^2 f$ is not available; the idea is then to take $d_k = -G_k \nabla f(x_k)$, where the matrices G_k are computed recursively, with the aim of approaching the inverse of $\nabla^2 f$ — just as x_k is supposed to approach a solution of (1). These are quasi-Newton methods, upmosty excellent, although little known out of the world of professional optimizers (and yet, they were discovered in 1959 by W.C. Davidon, a physicist!). They generalize to several dimensions the so-called secant method, or "regula falsi", which works as follows: to solve the nonlinear equation $g(x) = 0$ ($x \in \mathbb{R}$), the Newton's method is $x_{k+1} = x_k - g^{-1}(x_k)g(x_k)$. In the secant method, the unknown $g'(x_k)$ (playing the role of $\nabla^2 f(x_k)$) is approximated by the differential quotient $[g(x_k) - g(x_{k-1})] / [x_k - x_{k-1}]$ (playing the role of G_k^{-1}). In other words, the unknown but necessary differential information of g is identified upon observation of g -differences. In \mathbb{R}^n , where g is ∇f , the differences $y := g(x_+) - g(x)$

and $s := x_+ - x$ are stored. At iteration k , k such differences can be used to form an approximation G_k of $\nabla^{-2}f(x_k)$. Several such approximations are possible, one is admittedly best, usually called BFGS (for Broyden, Fletcher, Goldfarb, Shanno). At the first iteration, one starts with a starting approximation of $\nabla^{-2}f$. Then, having G_k and the current pair (s_k, y_k) of differences, G_{k+1} is explicitly computed by the formula (we drop the index k to alleviate notations)

$$G_+ = G - (sy^T G + Gys^T) / s^T y + (1 + y^T G y / s^T y) ss^T / s^T y. \quad (5)$$

Note that the whole information making up G_k is contained in k pairs of n -vectors (instead of n^2 numbers); this makes it possible not to store the matrices G_k themselves, but rather the pairs (s, y) ; economic formulae allow the computation of the direction without an explicit use of the recurrence formulae (5). This is especially handy for large problems, opening the way to limited-memory methods, where the maximal number of pairs (s, y) to be stored is imposed by the user, depending on n and the computer. The most economic such method is the conjugate gradient, in which the recurrence formulae reduce to

$$d_{k+1} = -\nabla f(x_{k+1}) + \alpha_k d_k.$$

When f is a sum of squares, the technique of Gauss-Newton can be used: G_k is the inverse of the matrix $\sum \nabla F_v(x_k) \nabla^T F_v(x_k)$. It differs from the true $\nabla^{-2}f$ by the term $\sum F_v \nabla^2 F_v$, which is small when the $\nabla^2 F_v$ are small (regression moderately nonlinear) or when the F_v are small (equations almost satisfied). We thus have one more way of approximating second order with the sole help of a first order simulator. In "bad" cases, the above G_k is only positive semi-definite; then, a term λI is added, so as to make it positive definite: this is the technique of Levenberg-Marquardt.

For the sake of completeness, we mention that the concept of second order is irrelevant for nondifferentiable problems (what is the second derivative of a function which has no first derivative?). In this case, the mere requirement $f(x_{k+1}) < f(x_k)$ is already difficult enough to achieve. There exist subgradient methods, in which this descent property is just given up, and bundle methods, highly sophisticated, in which computing the direction is a nontrivial problem "per se" (while explicit formulae give this direction in all cases above).

For linearly constrained problems, there are feasible directions methods with an active set strategy: one proceeds essentially as before, except that the direction is "twisted" so that $x_k + td_k$ satisfy the constraints for $t > 0$ small enough. In these methods, each x_k is feasible and φ can be chosen equal to f . Specifically, one chooses a set $I \subset \{1, \dots, m\}$ of constraints that one wants to keep active from x_k . Having the matrix G_k (whatever it is), d_k solves

$$\begin{cases} \min \nabla f(x_k) + \frac{1}{2} d^T G_k d \\ \nabla g_i(x_k) d = 0 \quad i \in I \quad \text{and} \quad \nabla g_{m+j}(x_k) d = 0 \quad j = 1, \dots, p \end{cases}$$

(note that this amounts to solving a linear system with $n+p+1$ equations). Then, a line-search is made according to the same principles as before, with the extra requirement that $x_k + t d_k$ must be kept feasible; in practice, this imposes a maximal value to t_k . A characteristic of these methods is that the set I of "activated" constraints can change by at most one index at each iteration. This can be a very inconvenient limitation if the number of linear constraints is large; then, it is often rewarding to pretend they are nonlinear and to place oneself in the next case.

Currently best methods for nonlinear constraints are based on the following principles: at each iteration, (1) is replaced by a "tangent linear-quadratic problem" at x_k . Indeed, d_k is the value $x - x_k$ obtained after solving (3). The iterates are not necessarily feasible, and the line-search is performed with the help of the function φ_r . The theory of Lagrange multipliers plays an important role; in particular, it is used to construct the matrix A in (3) and to select proper values of r . These methods are rather sophisticated; among others, solving (3) is by no means trivial, due to the presence of inequality constraints, even though linear.

V. EXAMPLE : IDENTIFYING A MOLECULAR STRUCTURE

An important problem in biochemistry is to figure out the location in the space of the atoms making up a given molecule. To do this, there are two possible approaches:

- Experimental methods. The actual molecule is observed by some physical means (X-ray diffraction, nuclear magnetic resonance,...) giving some information on the Fourier transform of its electron distribution.
- Theoretical methods. To each possible configuration is attached a potential energy, which has to be minimal at the actual configuration.

They both give birth to an optimization problem, and we will consider the second approach, a lot easier to explain. Let the atoms be numbered from 1 to N and characterize by $X_i \in \mathbb{R}^3$ the position of the atom i . Note: X_i can be thought of as the 3 Cartesian coordinates of the atom i , but it may be sounder to choose other parametrizations. A first important thing to understand is that our problem is solved if we can assign "the" correct values to X_1, X_2, \dots, X_N . In words, we have a problem with $n=3N$ unknowns. The second important thing is that, when these unknowns are assigned an arbitrary numerical value, a "performance index" (the conformational energy) can be computed, to tell us how good are these values.

This energy is the sum of a number of terms, for example

- Bond length: if d is the distance between two given atoms, there is an energy related to it, say

$$D(d) := \frac{1}{2} \delta (d - \tilde{d})^2$$

where δ and \tilde{d} are two constants characterizing the considered pair of atoms.

- Valence angle: if θ is the angle formed by 3 atoms, another energy is (β and $\tilde{\theta}$ are constants as before)

$$B(\theta) := \frac{1}{2} \beta (\theta - \tilde{\theta})^2.$$

- Van der Waals: again for a pair of atoms, another energy is

$$V(d) := v_1 (\bar{d}/d)^6 - v_2 (\bar{d}/d)^{12}.$$

We limit our description to these three terms, just for simplicity (needless to say, other terms can be considered as well without impairing the approach: electrostatic, torsion angles...). The total energy associated with a given configuration is then

$$\sum_{i,j \in \Delta} D[d(X_i, X_j)] + \sum_{i,j,k \in A} B[\theta(X_i, X_j, X_k)] + \sum_{i,j \in W} V[d(X_i, X_j)]$$

in which Δ , A and W are appropriate subsets of pairs and triples of atoms. Again, what is important is to understand that, in this expression, each individual term depends on the respective distance and angle of the relevant atoms; these in turn depend on the corresponding X -variables. Thus, we are entitled to call $f(x)$ the above expression, if we symbolize by $x \in \mathbb{R}^n$ the $n=3N$ unknowns.

It can be conceived that the computation of f and its derivatives can be programmed, provided that enough time is allotted to do a proper job – and debug it !. The result is a set of Fortran subroutines which, having numerical values of x as input, do all the necessary computations (geometric, trigonometric,...) and return the numerical values of f and ∇f , and perhaps $\nabla^2 f$. These computations are done even though the given x may represent an absurd configuration: they just simulate the proposed molecule (any absurdity will be reflected in the resulting f -value anyway; we pass under silence the question whether the above expressions are a reasonable model of the physical situation; quite another story). In a word, we are exactly in the situation of the tennis-practice alluded to in § III.

With the help of this program, a library code can be used to minimize f (or at least try to !). In applications of interest, the number N of atoms is in the 10^3 -range, so we are in the case 1.c of § II. It is advisable to choose a limited-memory algorithm, using only f and ∇f . We mention here that the sets Δ , A , W depend on x to some extent, which implies some complication: the human collaboration cannot be totally eliminated between the chemist, who programs the simulator, and the mathematician, responsible for the optimization code.

A specific software, called ORAL, has been written by K. Zimmermann, of Laboratoire de Pharmacologie Macromoléculaire, Institut Gustave Roussy (CNRS). The energy incorporates the AMBER force-field, and minimization is done by a limited-memory quasi-Newton code written at INRIA. It typically reaches RMS-gradients smaller than $10^{-2} \text{ kcal/mole/\AA}$ in a number of simulations substantially less than the number of atoms (when this number is in

the 10^3 -range). When confronted to similar problems, the original version of AMBER (based on the conjugate gradient method) simply does not reach such RMS-values.

For an illustration, we used a molecule having 1184 atoms (B-DNA d(A)₁₈.d(T)₁₈); it appeared to be fairly difficult to optimize. An unbounded cutoff was used (i.e. the sets like Δ , Λ and \mathcal{W} above were constantly maximal; otherwise AMBER had difficulties with discontinuous f). The evolution of the energy is plotted on Figure 1, as a function of the number of simulations for AMBER (dashed line) and ORAL. The left part displays the first 500 simulations (the starting energy was 1747 kcal/mole), and the tail appears on the right part, with a dilated vertical scale. After some 1400 simulations, the RMS values were .069 and .014 for AMBER and ORAL respectively. Note that comparison of speeds should be done on horizontal lines rather than vertical: for example, the level -700 was reached by AMBER in some 800 simulations, while 400 sufficed for ORAL.

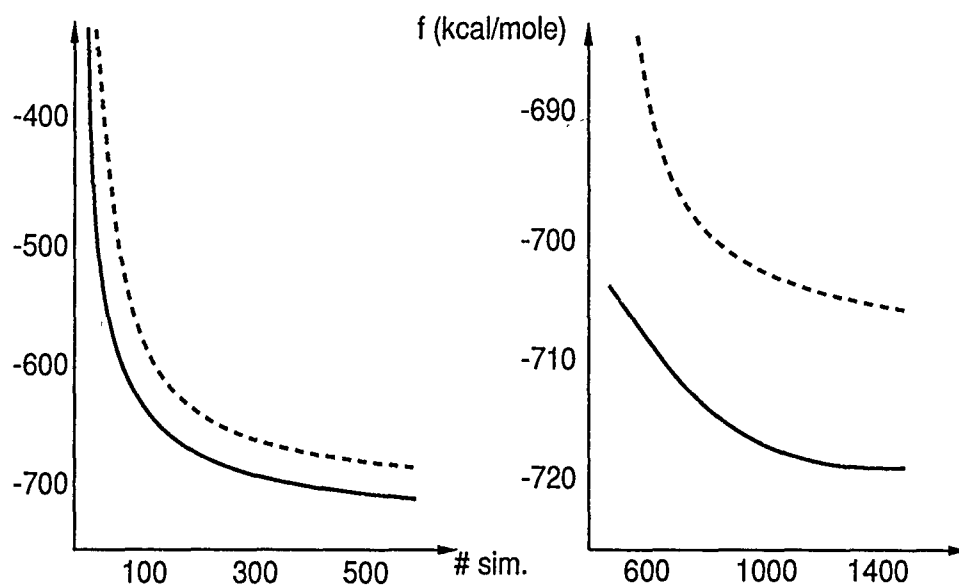


Figure 1.- Behaviour of AMBER and ORAL

VII. SHORT BIBLIOGRAPHY

For an introduction hardly more developed than the present text, see

C. Lemaréchal: Optimisation. in: *Précis d'Automatique*. Techniques de l'Ingénieur, 21 rue Cassette, 75006 Paris (1983).

For more complete texts: a good French book is

M. Minoux: *Programmation Mathématique*. Dunod, Paris (1983)

while many books are available in English; for example

R. Fletcher: *Practical Methods of Optimization*. Wiley, Chichester (2nd Edition, 1987).

P.E. Gill, W. Murray, M.H. Wright: *Practical Optimization*. Academic Press, New York (1981).

These two books are absolutely non-technical (theorems are rare, especially in the second). They contain many practical developments, useful for a user not particularly interested in technical details of optimization.

The best reference for unconstrained problems is probably

J.E. Dennis, R.B. Schnabel: *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Prentice Hall, Englewood Cliffs (1983).

Written by two world experts, it makes a good balance between mathematical rigour and readability. Furthermore, its content is of good quality, in that the study is limited to "good", i.e. efficient and modern, methods. Let us recall that methods for unconstrained problems form the basis for the constrained case.

Excellent references, with particular attention to constrained problems:

D.G. Luenberger: *Introduction to Linear and Nonlinear Programming*. Addison Wesley, Reading, Massachusetts (2nd Edition, 1984).

D.P. Bertsekas: *Constrained Optimization and Lagrange Multiplier Methods*. Academic Press, New York (1982).

Finally, we mention a popular but definitely outdated reference:

D.J. Wilde: *Optimum Seeking Methods*. Prentice Hall, Englewood Cliffs (1964).

DISCUSSION

KOZELKA - 1. We have used ORAL for energy minimization of cis-platinum adducts with oligonucleotides, as well as for calculations of energy barriers between different conformations. ORAL is really very robust, does not get stuck in high-energy states, and the energy gets smoothly and steadily down in the course of the minimization.

2. When comparing the efficiency of different minimizers, I would prefer to have the energy plotted against CPU time rather than against number of cycles. The CPU time used for one cycle can be quite different in different programs.

LEMARECHAL - Your point 2. is correct in general. More precisely, quasi-Newton algorithms with limited memory need a CPU time $T_A \approx O(n)$ at each cycle. It increases with the number m of pairs (s,y) alluded to in the paragraph following (5) in my paper ; this m , however, must be fairly small, say ≤ 20 : see numerical experiments in :

J.C. Gilbert, C. Lemaréchal : Some numerical experiments with variable storage quasi-Newton algorithms. Mathematical Programming 45, 3 (1989, to appear).

In the present class of problems, T_A is smaller than the time T_S spent in the simulator, unless an unrealistically severe cutoff is used to compute the force-field. For the case of Figure 1, there is no cutoff : $T_S = O(n^2)$, hence $T_S + T_A \approx T_S$ and the curves remain the same if the horizontal axis represents the total CPU time. With a reasonable cutoff, the ratio T_S/T_A would be about 10 : the number of cycles would still be a sensible measure of speed - and much more accessible than CPU !

CHOMILIER - The optimization algorithm is able to refine by iteration the value of the minimum, i.e. to give x_{k+1} starting from x_k .

What about the first point of the problem x_0 . How can it be determined ? Is there any algorithm to compute it ?

LEMARECHAL - Unfortunately, the answer is a definite "no". There are problems with a special structure, well-suited to special mathematical tools : convex analysis, duality theory (this is not the case of the present problems, I believe) ; then a solution can be grossly approximated, say by x_0 . Otherwise, no answer can be given to these questions, at least by numerical analysts.

STONE - Can you highlight the differences between ORAL and the routines available in the NAG and Harwell libraries?

And can you comment on the problem of escaping from local minima when the objective is to find the global minimum?

LEMARECHAL - The code is probably comparable to E04DGF of NAG. Technically, the mathematical method is that of VA13A (Harwell). The line-search is similar and the direction is computed by a quasi-Newton formula such as (5). However, the G-matrix is not quite the same: in Harwell, it is computed on the basis of all the pairs (s_1, y_1) , (s_2, y_2) , ..., (s_k, y_k) and is stored explicitly; so the program needs $O(n^2)$ places of memory. In the present program, the matrix uses only (s_{k-m}, y_{k-m}) , ..., (s_k, y_k) , with m small (see my answer to J. Kozelka above); thus, only $O(n)$ memory is needed, say $30n$. This way of computing the direction is described in

J. Nocedal: Updating quasi-Newton matrices with limited storage. Mathematics of Computation 35(1980) 773-782.

The question of local vs. global optima is very close to that of J. Chomilier above, and the answer is unfortunately the same: we have no practical tool for (i) eliminating local minima, or (ii) checking global optimality. To realize that the existing tools are fairly weak, see

L.C.W. Dixon, G.P. Szegö: Towards global optimization, vols. I (1975) & II (1978), North Holland.

On the other hand, the main trouble seems to be that the optimization codes provide you with stationary points (where the gradient is 0) rather than true local minima (where also the Hessian matrix is (semi) positive definite). Incidentally, so is probably the case of Figure 1. Now, there exist methods - rather recent and still rather experimental - which are able to avoid local maxima and saddle-points, hence finding true local minima (see below my answer to J. Devillers).

SOUMPASIS - Simulated annealing is a positive approach, which "globalizes" local methods via Monte Carlo techniques. To my knowledge, it has not been applied to molecular reconstruction (except perhaps in crystallography) but good success has been reported in combinatorics, for example for the travelling salesman problem.

LEMARECHAL - I have some misgivings concerning simulated annealing in the present context: the objective function must be very simple, because it is going to be computed millions of times. So is not the case of the force-field, unless a concept of "neighboring" configurations is defined, to allow fast computation of f , given its value at some neighbour.

PETTIT (comment) - Truncated-Newton and quasi-Newton methods are reminiscent of the adopted basis Newton-Raphson method where the last m gradient vectors are used to define a small subspace of most gain. Then a full N-R step is made in this space of dimension 5-10. This method works well for problems of dimension up to 20.000 or 30.000.

DURUP - For the dialogue between physical chemists and numericists it should be pointed out that the choice of the best algorithm is dependent on (i) whether we are interested in the absolute minimum or in a local minimum, and (ii) the topography of the potential surface. The mathematician should not "ignore" $f(x)$. Moreover, a common language should be developed for describing the topological features which are critical for the definition of the algorithm, which for example will not be the same depending on whether there are 1, 1000, 10^{10} or zero paths going (monotonously decreasing) from $f(x_0)$ to the minimum; and what the "widths" of these paths in the \mathbb{R}^n space are.

LEMARECHAL - The mutual "ignorance" between the algorithm and the simulator, which I believe is necessary, concerns only the Fortran run-time. It excludes by no means a close collaboration between the algorithm-designer and, say, the physical chemist, when necessary ; for example during the phase of choosing an adequate algorithm : the computer is not involved, yet.

As for your second point, we numerical analysts are disarmed to study the topology of the energy-surface. I would rather say that this kind of question is in the field of pure mathematics but, with 10^4 variables, I am not very optimistic. The issue would probably imply a proper parametrization of these variables, so as to reduce their number.

DEVILLERS - Concerning molecular mechanics, second derivatives are rather easy to calculate. So, provided that the problem is not too large, in what consists the superiority of Davidson-Fletcher-Powell type of optimization methods versus Newton-Raphson type ?

LEMARECHAL - Newton-Raphson (N-R) methods have the definite advantage of converging much faster : instead of the 10^3 cycles reported on Figure 1, 10^2 would already reduce the RMS to much lower values. The price is the need to solve the linear system (4). Also, the method rushes (when it works) to the next stationary point, which

most probably is not a local minimum.

I believe, however, that N-R methods are best for the present problem, and that their disadvantages can be turned to advantage over quasi-Newton methods. Indeed, (4) equates to 0 the gradient of the quadratic approximation - call it f_q - of f near x_k ; this is unwise if f_q is not convex. A much better idea is to minimize f_q by some economic method such as conjugate gradient, and only approximately. In particular, such a local minimization is interrupted when the trial direction is downhill and has negative curvature (thus indicating that f_q is unbounded from below, and that N-R is not appropriate). This approach is described in :

R.S. Dembo, T. Steihaug : Truncated-Newton algorithms for large scale unconstrained optimization. Mathematical Programming 26,2 (1983) 190-212.

In addition to being implementable for large-scale optimization, it avoids saddle-points of f . With J. Navaza (Fac. Pharmacie, Univ. Paris-Sud), we have already obtained encouraging results for the phase-problem in crystallography ; it should certainly be tried for energy-minimization as well.

Now, when the problem is not too large, (4) can be solved explicitly but there remains the difficulty of negative eigenvalues. Against it, one can minimize f_q on a ball of radius $t > 0$, centered at x_k . Even though f_q is not convex, its global minimum can then be computed. Furthermore, the line-search is replaced by an adjustment of t at each iteration k . We obtain a "trust-region" algorithm, which captures all the advantages of N-R, and which converges to a true local minimum. See a review in :

J. Moré : Recent developments in algorithms and software for trust region methods. In A. Bachem, M. Grötschel, B. Korte (Eds.) Mathematical Programming, the State of the Art ; Springer Verlag (1983) 258-287.

DESCRIBING PROTEIN CONFORMATION : A NEW MATHEMATICAL APPROACH

K. ZAKRZEWSKA and R. LAVERY

Laboratoire de Biochimie Théorique, Institut de Biologie Physico-
Chimique, 13, rue Pierre et Marie Curie, Paris 75005 (France)

ABSTRACT

An algorithm, P-Curves, recently developed in our laboratory enables a generalized helicoidal description of protein structure and yields both an overall axis describing the folding of the protein backbone and a full set of helicoidal parameters locating each peptide in space. Two applications of this algorithm are presented. Firstly, P-curves is used to automatically detect secondary structures within proteins, the definition of the secondary structure limits being determined statistically with respect to random polypeptide chain conformations. Secondly, the method is used to search and quantify conformational similarities between homologous proteins.

INTRODUCTION

Although a considerable number of protein structures have now been crystallographically determined to high resolution there are still difficulties in extracting all the conformational data that these results contain. In particular, there is no rigorous procedure for the precise location of secondary structures and the description of their deformations, or for determining the exact pathway followed by a folded polypeptide backbone.

We have thus been led to develop a rigorous algorithm for dealing with such problems. Our starting point was to look for a way to extend helical geometry to irregular systems so that the notion of a helical axis could be conserved, although this axis would in general be a space curve rather than a straight line. The first such an algorithm was developed in our laboratory for analysing the irregular conformation of nucleic acids (refs.1-2), and soon afterwards the necessary adaptations needed to treat proteins were made (ref.3) yielding the algorithm named P-Curves.

P-Curves has obvious applications for describing protein

folding patterns and for the automatic comparison of related proteins. It can be applied to any protein for which the atomic coordinates are known, from crystallographic data banks or from molecular mechanics or molecular dynamics studies. In this paper we will describe two applications of the method, firstly, an automatic detection of secondary structures and secondly, an analysis of fine differences between homologous protein conformations.

P-CURVES METHODOLOGY

We start by presenting briefly the P-Curves algorithm. The full mathematical details of which can be found in our earlier publications (refs.1-3). The most fundamental aspect of the algorithm we have developed concerns the definition of an axis, or, more precisely, a space curve, which will describe the way the molecule studied is folded in 3 dimensions. In the case of a perfect helix, this axis is a rigorously defined straight line, termed the helical axis. Every monomer has the same position and orientation with respect to its local axis segment and each monomer in the structure can be reached from the preceding monomer by a fixed rotation coupled with a fixed translation along the axis.

P-Curves approach is a natural extension of this idea to the case of irregular systems. Its basis is the definition of a function which describes departure from perfect helical symmetry in terms of the curvature and dislocation of the axis describing the polymer and in terms of changes in the position of successive monomers with respect to this axis. Minimisation of this function yields the optimal space curve describing of the polymer where both types of irregularity have been smoothed in a least-squares sense. Since the function is constructed so as to take into account simultaneously the position of all the monomeric peptide units making up the polymer, the final description of any one of these units depends on the position of its neighbors. This leads to a much more coherent view of the overall conformation than that obtained with any purely local parameters such as the backbone torsion angles.

After optimisation the irregularity of each pair of adjacent peptides within the structure can be measured, to yield a value

called the dimeric irregularity function or 'DIF'. The DIF function is particularly helpful in locating protein secondary structure zones without having to make reference to local properties such as ϕ/ψ angles or hydrogen bond geometries.

APPLICATIONS

The first application of P-Curves methodology concerns the automatic detection of secondary structure zones within proteins. In a regular helix ("helix" in this sense includes also ribbon and sheet conformations) the DIF values are zero. In order to detect secondary structures in real proteins, some degree of irregularity has to be accepted. The limiting DIF values for segments of different length were consequently determined with respect to random conformation polypeptide chains. To this end 26 proteins containing a total of roughly 3000 peptides, resolved to at least 1.8Å, were selected from the Brookhaven Protein Data Bank (ref. 4). These proteins were analysed for the proportion of glycine and proline. 20 random conformation polypeptides were then constructed with lengths varying between 100-400 peptides, by sampling ϕ and ψ angles from the accessible regions of the Ramachandran plot (refs. 5,6). Glycine and proline residues were located randomly within these polypeptides in a proportion corresponding to that found in the real proteins and their conformational angles were sampled from the ϕ/ψ maps correctly describing their respective allowed zones.

The random polypeptides created in this way were subsequently analysed by the P-Curves program and maximum DIF values were calculated over all fragments with lengths from 1 to 30 peptides long. In Fig 1 we give several examples of the distributions obtained for the fragment lengths 4,8,12, and 16 peptides. (the abscissa gives the maximum fragment DIF values and the ordinate shows their relative frequency). In Fig 2 the same distributions are shown for our set of well-resolved proteins. Very high frequencies can be seen in the region of small DIF values, corresponding effectively to secondary structure zones. From these distributions we were able to calculate the maximal DIF value which would correspond to the existence of secondary structure. Initial trials indicated that a value corresponding to 5% random probability was the most appropriate choice.

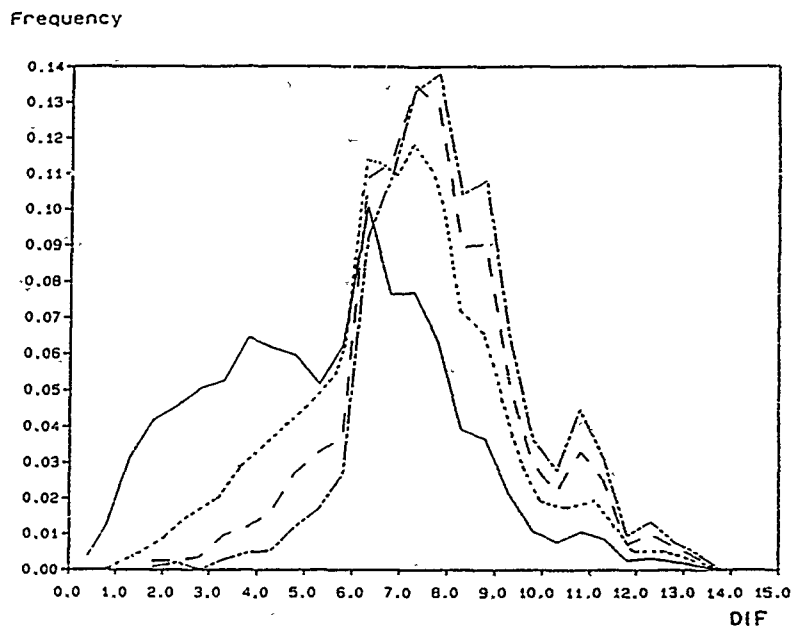


Figure 1. DIF distributions for random polypeptides. Distributions are shown for fragments of length: 4,8,12,16 peptides.

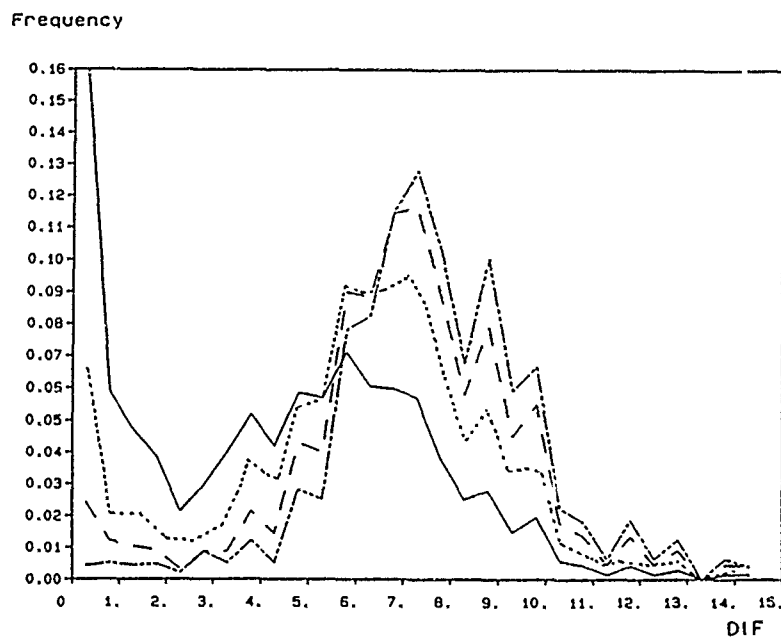


Figure 2. DIF distributions for true proteins. Distributions are shown for fragments of length: 4,8,12,16 peptides.

Having determined a set of limiting DIF values for each fragment length it is now possible to automatically detect secondary structure regions. In order to distinguish between the different types of structure which may occur we simply inspect the helicoidal parameters of the peptides in each zone. We have shown (table II, ref.3) that, with the exception of Xdisp, all the inter-peptide and peptide-axis parameters can be used to identify secondary structure types. As an example, Fig. 3 shows the distribution of Ydisp in our protein set. Two distinct peaks are present, the first one centered at the origin corresponds to beta sheets, while the second at 1.5 Å are α helices. The small shoulder at 1.1 Å corresponds to 3/10 helices.

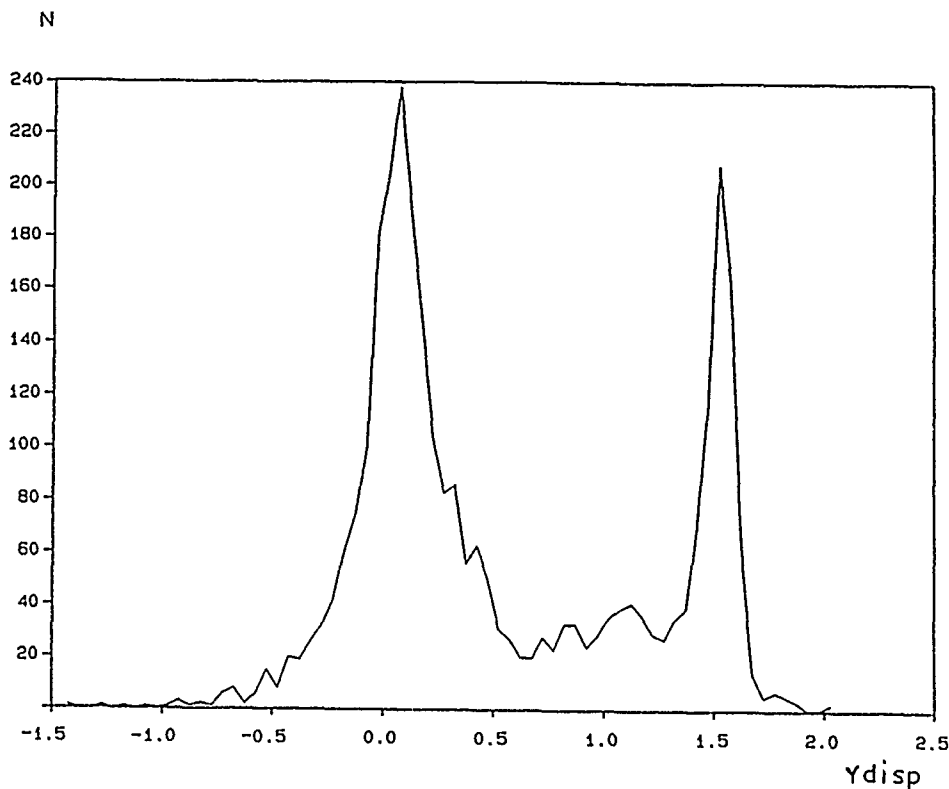


Figure 3. Ydisp distribution for the chosen set of well resolved proteins.

The application of P-Curves to the automatic detection of the secondary structures is illustrated by two specific proteins. The first one is erythrocrucorin from chironomus thummi thummi whose "fingerprint" of DIF values of all successive dipeptides is shown in fig 4. This figure indicates 6 regions of secondary structure,

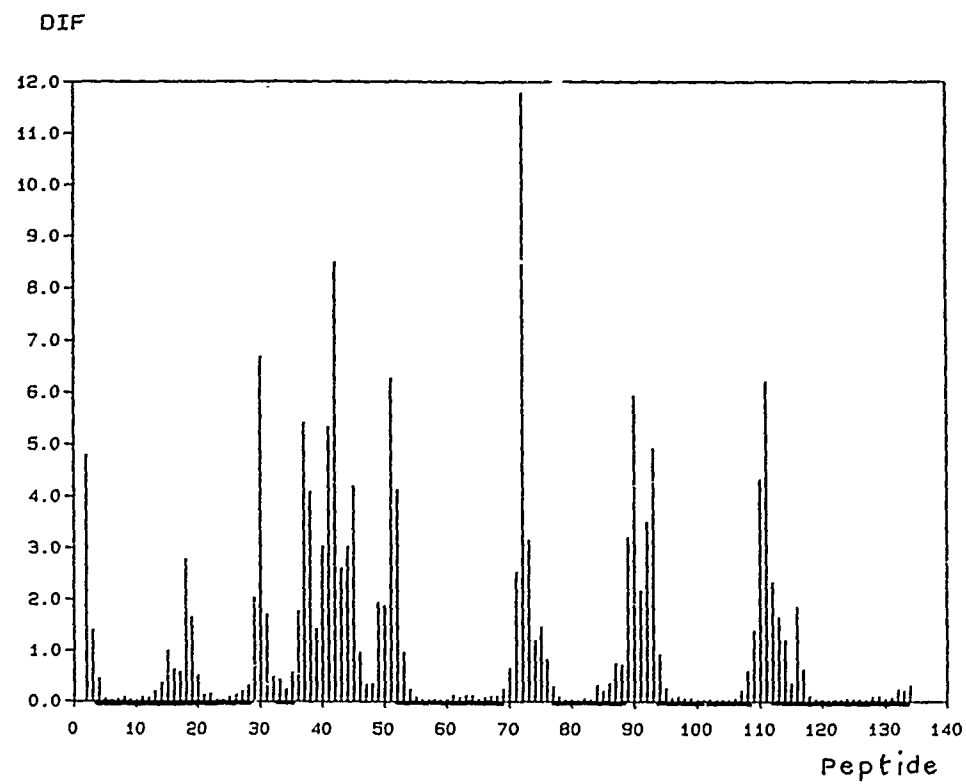


Figure 4. DIF values for erythrocrucorin.

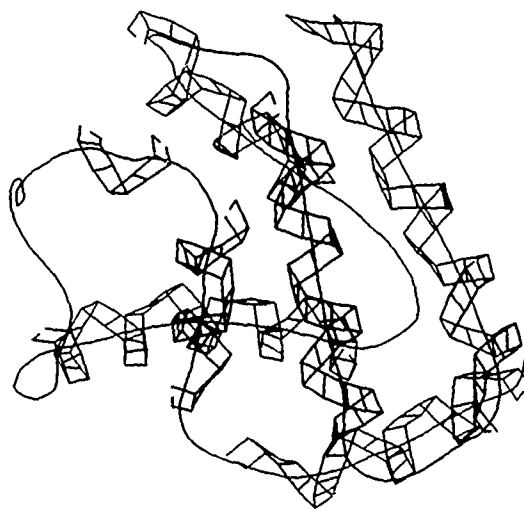


Figure 5. Secondary structures within erythrocrucorin.

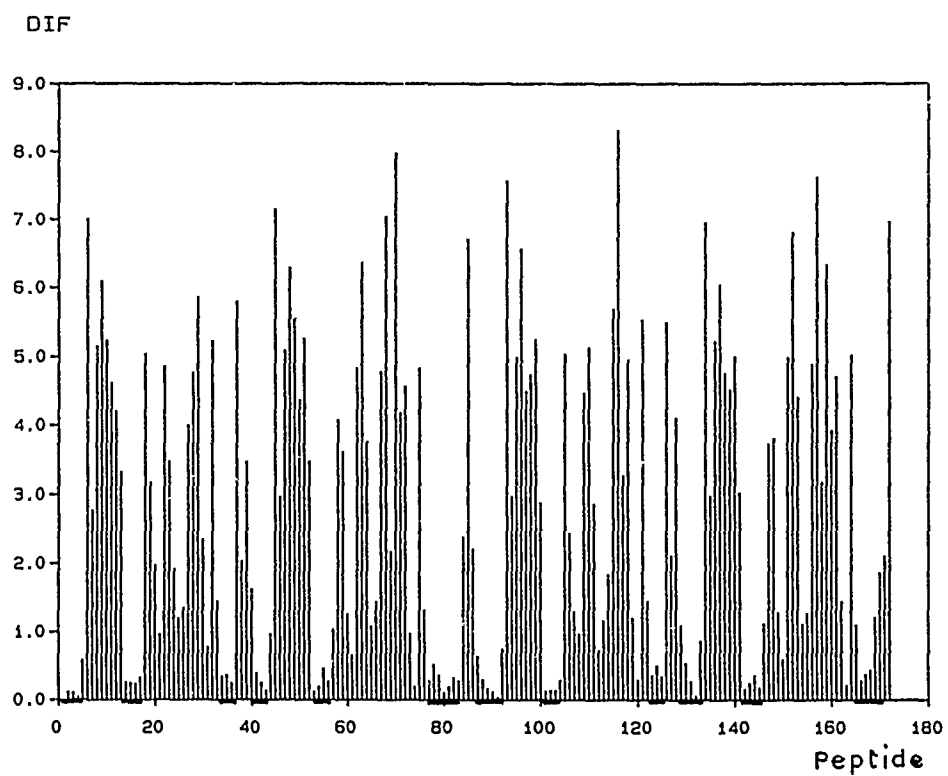


Figure 6. DIF values for crystallin.

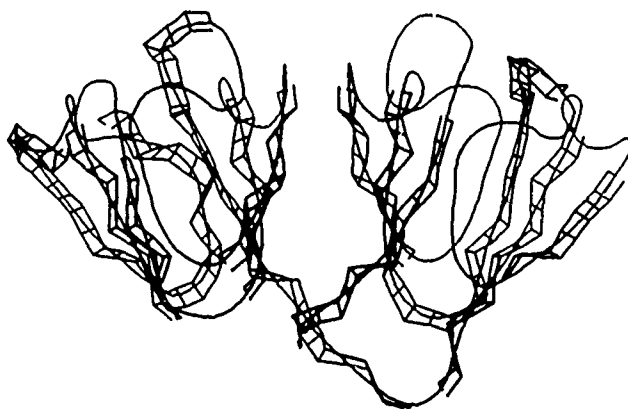


Figure 7. Secondary structures within crystallin.

which were all identified as α helices. Fig 5 shows our so-called "ribbon representation" (ref. 3) of the protein structure. The curved line represents the generalised helical axis characterising its folding while the ribbon shows the position of the polypeptide backbone. In order to make the secondary structures more visible the ribbon corresponding to irregular segments has been omitted. The second protein presented is crystallin from the bovine eye lens. Its fingerprint (fig 6) shows a heavy domination of irregular structure. 12 short β sheets were nevertheless located. The corresponding ribbon diagram is given in fig 7.

DIF

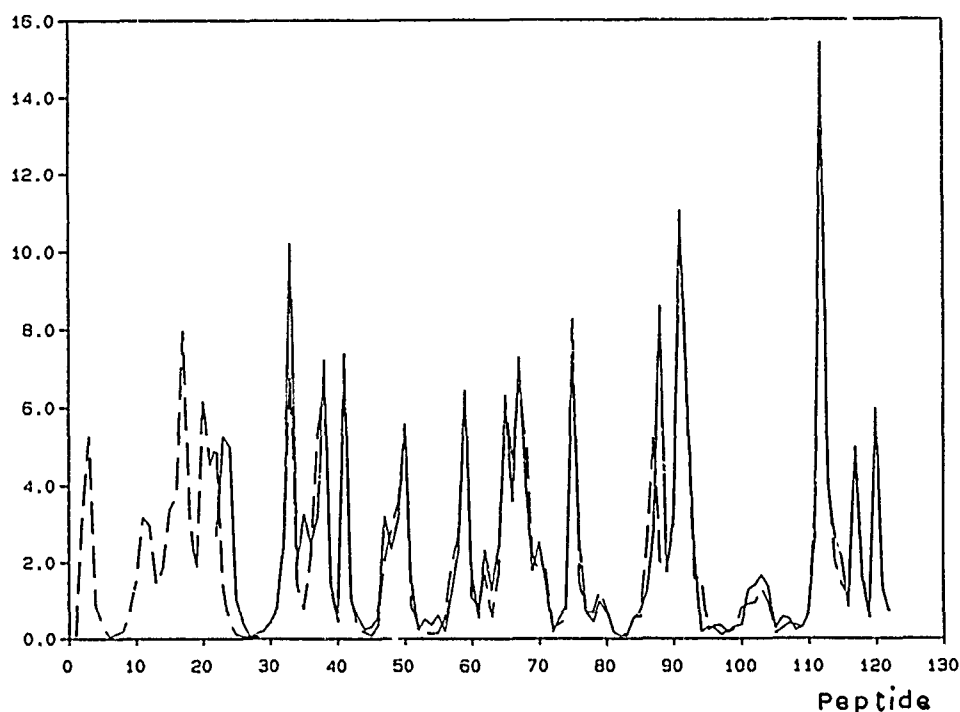


Figure 8. Superposition of DIF values for ribonuclease A (dashed line) and S-protein (full line).

The other application of P-curves that we will briefly describe concerns the detection of fine conformational changes within proteins. The example we present concerns the effect of cleaving the S-peptide (the first 20 residues) from ribonuclease A. In fig 8, the fingerprints for both the cleaved and the complete structures are superposed, the full line corresponding to the S-protein, the dashed line to the full protein. It can

immediately be seen that these two molecules yield almost identical results. The only differences can be found in the vicinity of peptide 35 and in the region of the peptides 85-90. The similarity of the two structures is confirmed in Fig 9, where for clarity only the helical axis is given. The two conformations are effectively seen to be similar, however, one difference, consisting of a hinge motion opening outwards the loops on the right hand side of the figure, is visible. We can thus see that P-Curves analysis and, in particular, the DIF "fingerprint" is a powerful way of localising and quantifying global conformational changes.

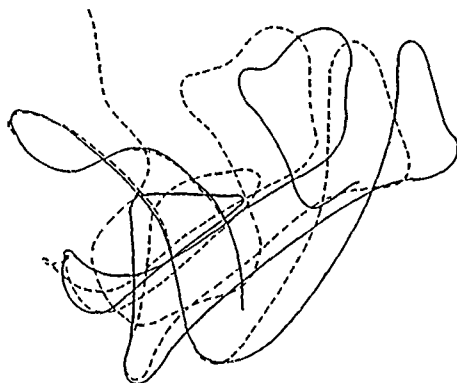


Figure 9. Helical axis for Ribonuclease A (dashed line) and S-protein (full line).

CONCLUSIONS

We have demonstrated that the P-Curves algorithm can be used to quantitatively describe protein conformation. The algorithm leads to a full set of independent, commuting parameters and, most importantly, to a global axis which concisely describes the macromolecular conformation. An automatic detection of the secondary structure based on precise, statistically defined rules, has also been developed. Overall, the method is useful for comparing the structures of different proteins and for detecting and defining conformational changes upon, for example, substrate binding. It is also certainly of interest for interpreting dynamic simulations of proteins, where the extraction of clear conformational data is often a major problem.

P-Curves is currently being extended to the analysis of irregular zones such as the beta turns (ref.7) and to the

definition of the spatial arrangement of secondary structure motifs within proteins.

REFERENCES

- (1) R. Lavery and H. Sklenar, The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids, *J. Biomol. Struct. Dynam.*, 6 (1988) 63-91.
- (2) R. Lavery and H. Sklenar, Defining the structure of irregular nucleic acids : conventions and principles, *J. Biomol. Struct. Dynam.*, 6 (1989) 655-667.
- (3) H. Sklenar, C. Etchebest and R. Lavery, Describing protein structure : a general algorithm yielding complete helicoidal parameters and a unique overall axis, *Proteins : Structure, Function and Genetics*, in press.
- (4) F.C. Bernstein, T.F. Koetzle, G.J.B. Williams, E.F. Meyer, M.D. Brice, J.R. Rogers, O. Kennard, T. Shimanouchi and M. Tasumi, The protein data bank : A computer-bases archival file for macromolecular structures, *J.Mol.Biol.*, 112 (1977) 535-542.
- (5) C. Ramakrishnan and G.N. Ramachandran, Stereochemical criteria for polypeptide and protein chain conformation II. Allowed conformations for a pair of peptide units, *Biophys. J.*, 5 (1965) 909-933.
- (6) G.N. Ramachandran and V. Sasisekharan V, Conformation of polypeptides and proteins, *Adv. Protein Chem.*, 23 (1983) 283-437.
- (7) P.Y. Chou and G.D. Fasman, β -Turns in proteins, *J. Mol. Biol.*, 115 (1977) 135-175.

DISCUSSION

DURUP - 1) How, i.e. according to what kind of sampling, do you define what you used as "random polypeptides" ?

2) Did you develop a method for describing the relative positions of two helices, i.e. the most important parameters characterizing a loop ?

ZAKRZEWSKA - 1) The random polypeptides were constructed by sampling the phi/psi torsion angles from the allowed regions of the Ramachandran map. Taking into account the nature of the peptide involved glycine, proline or others.

2) I am working on this problem at the moment.

GENMOL

A FAST PROGRAM FOR MOLECULAR MODELING. APPLICATION TO THE
DETERMINATION OF THE PSYCHOTONIC OR SEDATIVE EFFECT OF TRICYCLIC
ANTIDEPRESSANT DRUGS

Gérard PEPE and Didier SIRI

Centre de Recherche sur les Mécanismes de la Croissance Cristalline*,
Campus de Luminy, Case 913, 13288 Marseille cedex 9 (France)

*Laboratoire propre du C.N.R.S. associé aux Universités d'Aix-Marseille III.

SUMMARY

GENMOL is an interactive program of force field calculation written in Fortran 77 (~12000 lines) running on any 32 bits computer equipped with a graphic terminal, or a graphic station under PHIGS. The program is designed to build any molecule from a few atoms to hundreds of atoms. Its fastness (10 to 100 times faster than MM2) allows one to perform conformational analysis at each step of the molecular building. When refining the program considers the total strain energy and the local strain energies. The π systems are taken into account. The preferred conformations of macrocycles are generated by a special algorithm. In this work an application of the program is given with the purpose to decipher the origin of psychotomic or sedative activity of tricyclic antidepressant drugs from a conformational analysis.

INTRODUCTION

The X-Ray diffraction is the standard method to solve the 3D structure of molecular compounds. Unfortunately the technique is very heavy and time consuming, and not always conclusive. GENMOL was initially designed as an alternative to yield specific structures which are expected to be close to structure variants reported yet.

The program might be able to keep rigid the common part of the molecules and to perform conformational analysis on the varying part of

the molecule. Further developments of the program led us to model π systems as to find the most probable conformation of macrocycles.

In a previous work (ref. 1) on tricycle neuroactive compounds (Neuroleptics and antidepressants), we pointed out the importance of the conformational role during the recognition step between the drug and the neurotransmitter receptor.

The results of the calculations performed on X-Ray structures led us to admit a lock and key interaction model, which means that the drug molecules interact with the receptor without conformational fit. As antidepressant drugs exhibit psychotonic or sedative effect in relation with their interaction with noradrenaline or serotonin receptors, in the lock and key model the preferred conformation of the psychotonic drug must fit to noradrenaline receptors while the preferred one of sedative drugs must do the same with serotonin receptors.

In order to gain some insight into this reaction mechanism, GENMOL was used to compute and to compare the preferred conformations of the antidepressant drugs (ref. 2) and those of the corresponding neurotransmitters.

THE PROGRAM

The program is designed to build or to modify molecules, several libraries of molecular fragments and of molecules are available.

The principal feature of the program is the concept of pivots which are all the interatomic bonds of linear chains and all the axis passing through all the non adjacent atoms of the cycles. Molecular deformations are done by rotation around the pivots, when running GENMOL in the pivot option, bond lengths are kept fixed. The program can perform a complete Force Field Calculation if needed. Parameters for stretching and bending potential functions are derived from MM2 (ref. 3). Those for van der Waals and Hydrogen bond potentials are issued from ECEPP (ref. 4). Electrostatic interactives are calculated in a monopole approximation with net atomic charges computed with :

- 1) the DEL RE'S method (ref. 5) for the σ part with a new set of atomic parameters (réf. 6).
- 2) An empirical approximation for the π part giving equivalent charges to PARISER and PARR ones (réf. 7).

Particularities of GENMOL

*Intelligence :

- Defines the hierarchy of rotations.

- Finds the nature of each bond :

(4 types of bonds : single bond, double bond, linearly conjugated double bond ($X=X...X=X$), aromatic and triple bond) in order to allow deformation of π systems.

- Gives the preferred conformations of a molecule, taking into account the cycle conformations.

- Corrects automatically the input errors with warning messages.

*Fastness : (10 to 100 times more rapid than MM2) necessary to perform conformational analysis.

Automated to build peptides or nucleic acids (DNA and RNA).

* Technical characteristics :

. Language : Fortran 77.

. Size : ~12000 lines.

. Computer : Any 32 bits computer equipped with a graphic terminal or any graphic station under PHIGS.

. Size of built molecules : 3 atoms to hundreds of atoms.

APPLICATION TO ANTIDEPRESSANT DRUGS AND RESULTS :

Antidepressant drugs block preferentially the presynaptic receptors of noradrenaline and serotonin (5HT for 5-Hydroxytryptamine). From a statistical point of view, psychotonics are secondary amines acting on noradrenergic receptors while sedatives are tertiary amines acting on serotonergic receptors (ref. 8). Let us notice that amine substitution has both conformational and electronic effects.

The conformations of 19 antidepressants were generated (the conformations with a differential strain energy ΔES lower than 10 Kcal mole⁻¹ were kept in the analysis, ΔES being the difference of strain energy to the most stable conformation) and then compared to the preferred conformations of the neurotransmitters. The total numbers of conformation molecules here studied and their basic structure origin (from X-Ray or calculated) are gathered in Table 1).

Table 1.

Antidepressant drugs and numbers of conformations analysed (XR for structures coming from X-Ray and CAL for calculated ones) in a scale going from psychotonic to sedative effect.

Compound	Symbol	N	Origin of the struc.	Ref.	Total nu. of conform. with Es<10 (Kcal/mole)
Noradrenaline	NAD	-	XR	9	7
Serotonin	SER	-	XR	10	8
Histamine	HIS	-	XR	11	9
Amineptine	AMI	1	XR	12	3
Metapramine	MET	2	CAL		1
Quinupramine	QUI	3	XR	13	1
Desipramine	DES	4	CAL		13
Protriptyline	PRO	5	CAL		14
Nortriptyline	NOR	6	CAL		15
Imipramine	IMI	7	XR	14	38
Clomipramine	CLO	8	XR	15	24
Demexiptiline	DEM	9	CAL		27
Propizepine	PRP	10	CAL		3
Dosulepine	DOS	11	CAL		8
Amoxapine	AMO	12	CAL		3
Butriptyline	BUT	13	CAL		16
Doxepine	DOX	14	CAL		10
Opipramol	OPI	15	CAL		59
Noxiptiline	NOX	16	XR	16	20
Amitriptyline	AMT	17	CAL		13
Maprotiline	MAP	18	CAL		10
Trimipramine	TRI	19	CAL		5

From BOLTZMAN statistics in the accuracy limits of our calculations, and at human body temperature, we can anticipate the presence in the synapse of only one conformation for noradrenaline (ND1) and of three conformations for 5-HT (SR1, SR2, SR3), the differential strain energies for SR2 and SR3 being respectively 0.5 and 1.5 Kcal mole⁻¹.

Table 2 gives the conformation numbers and the corresponding strain energies of the drug that superimpose with the preferred conformation of the neurotransmitters (the number allows to identify the conformation in order to see if it is the same that interacts with the different neurotransmitter receptors). The superimpositions with SR2 are absent because most of them have a differential strain energy greater than zero.

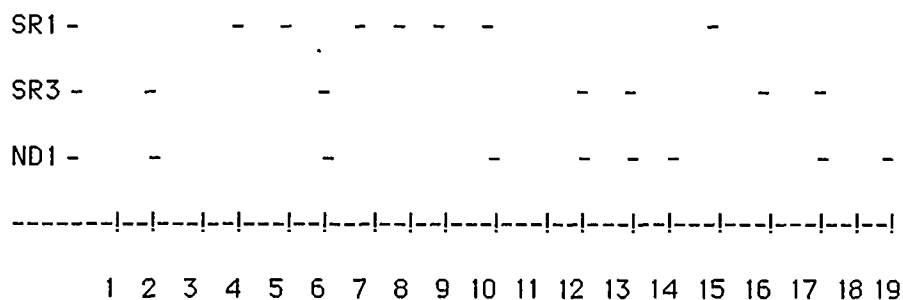
Table 2 - The best superimpositions between drug conformations and the preferred conformation of a noradrenaline and 5-HT present in the synapse.

Drugs		Superimpositions					
Symbol	num.	with ND1		with SR1		with SR3	
		conf. number	strain energy	conf. number	strain energy	conf. number	strain energy
AMI	1	-	-	-	-	-	-
MET	2	2	0.000	2	0.000	-	-
QUI	3	-	-	-	-	-	-
DES	4	11	1.70	11	1.70	10	0.00
PRO	5	-	-	2	2.80	8	0.80
NOR	6	14	0.00	14	0.00	13	0.40
IMI	7	3	9.40	8	1.50	12	0.00
CLM	8	16	6.10	16	6.10	24	0.00
DEM	9	9	3.10	9	3.10	27	0.00
PRP	10	3	0.00	-	-	3	0.00
DOS	11	8	0.40	-	-	-	-
AMO	12	1	0.00	1	0.00	-	-
BUT	13	16	0.00	16	0.00	7	4.70
DOX	14	10	0.00	-	-	3	0.70
OPI	15	3	6.90	3	6.90	38	0.00
NOX	16	13	8.50	5	0.00	8	1.50
AMT	17	6	0.00	6	0.00	9	1.60
MAP	18	-	-	-	-	5	1.60
TRI	19	4	0.00	-	-	3	1.80

The significant results of this analysis are displayed in figure 1, where the superimposition with the neuromediator conformations (ND1, SR1 and SR3) is sketched by a dash. The strain energy value of the drug conformation is zero or very close to zero.

Figure 1.

Diagram of superimpositions between the most stable conformation of the antidepressants and the neurotransmitter conformations present in the synapse supporting the possibility of the compounds to react with the corresponding receptors without conformational fit.



CONCLUSION ON THE APPLICATION :

The calculation of the conformations of the neurotransmitters and of the tricyclic antidepressant drugs with GENMOL and their comparison, allow to draw some meaningful conclusions regarding interaction of the antidepressants on the neurotransmitter receptors.

. Drug molecules and neurotransmitters generally interact in their preferred conformations (without conformational fit) with the receptors.

. The serotonergic receptors belong to two structurally different families, while there is only one structural family for the noradrenergic receptors. The psychotonic drugs interact on one family of these receptors and do not interact on the noradrenaline receptors, while the sedative drugs interact on the other family of 5-HT receptors and on the noradrenaline receptors.

REFERENCES

- 1 G. Pèpe, J.P. Reboul, B. Cristau and Y. Oddon, Eur. J. Med. Chem., 22 (1986) 339
- 2 G. Pèpe, J.P. Reboul and Y. Oddon, Eur. J. Med. Chem., 24 (1989) 1.
- 3 N.L. Allinger, Quant. Chem. Prog. Exc., MMP2, (1985), Bloomington, Indiana.
- 4 H.A. Sheraga, Quant. Chem. Prog. Exc., ECEPP, (1982), Bloomington, Indiana.
- 5 G. Del Re, J. Chem. Soc., 40 (1958) 31.
- 6 G. Pèpe, B. Serres, D. Laporte, G. Del Re and C. Minichino, J. Theor. Biol., 115 (1985) 571.
- 7 R. Pariser and R.G. Parr, J. Chem. Phys., 21 (1953) 446, 767.
- 8 D.J. Hlasta, D.L.R. Haubrich and D.L. Luttinger, CRC Handbook of CNS agents and local anesthetics, M. Vederame M. Florida, 1986.
- 9 D. Calstroem and R. Begin, Acta Cryst., 23 (1967) 313.
- 10 V. Thewald and C.E. Bugg, Acta Cryst., B 28 (1974) 2884.
- 11 K. Prout and S.R. Critchley, Acta Cryst., B 30 (1972) 82.
- 12 J.P. Reboul, B. Cristau and G. Pèpe, Acta Cryst., B 38 (1982) 1489.
- 13 J.P. Reboul, J.C. Soyfer, B. Cristau, C. Caranoni and G. Pèpe, Acta Cryst., B 38 (1982) 2633.
- 14 E.F. Paulus, Acta Cryst., B 34 (1978) 1942.
- 15 M.L. Post and H. S. Horn, Acta Cryst., B33 (1977) 2590.
- 16 G. Bandolini and M. Nicolini, J. Cryst. Spect. Res., 13 (1983) 191.

DISCUSSION

DEVILLERS - 1° Concerning molecules having delocalized π electrons, does your program perform quantum mechanical treatment before the mechanical one ?

2° I have been very impressed by the speed you mentioned for your program : up to 100 times faster than Molecular Mechanics 2. MM2 is already known to be fast because of the block-diagonal Newton-Raphson method used. So it brings me to ask you if you refine the geometry by changing all the coordinates of every atom at each step ?

PEPE - 1° The program only performs force field calculations.

2° The program uses non classical method to get the best conformations.

. It goes in the direction of the best conformations from topological informations.

. Then when the preferred conformation is found, it is optimized only by considering the non bonded interactions plus the torsional energy by rotating around bonds. The program minimizes all the local energies and computes the total strain energy only if deformations are significant.

DURUP - Just a warning. Since the binding of the messenger to its target aims at inducing some process in the target, many biochemists consider that this binding may induce some strain, which will be released mainly in the transition state under consideration. Therefore such possibilities should be tested by the model calculation.

PEPE - In my model, I only consider that a conformational fit will increase the activation energy which will slow down the rate of interaction.

KOZELKA - In your slides you have shown that a conformation of generated by Genmol fits the X-ray structure of the molecule, whereas a conformation coming from force-field calculation does not. Does this not simply reflect the facts that

- i) the X-ray structure is one of those the "mean value" of which was used by Genmol to define the tricyclic part of the molecule ;
- ii) the force-field calculations did not contain crystal packing forces ?

PEPE - The example given is to indicate that we must be careful about geometry results coming from force field calculations. Generally packing interactions are not strong enough to have any influence on the geometry of rigid parts of a molecule, that is the case of the tricycles of the antidepressant drugs.

SURCOUF - 1° In your examples, you have long chains ; you have performed conformational analysis. How do you select all the conformations with low energy value ?

2° Do you perform energy minimization before you select your conformations of lowest energy ?

PEPE 1° The program generates all the conformations of a molecule in a range of given strain energy, most of the wrong conformations are eliminated from topological conformations.

2° In the first step of calculations we select the regions of the preferred conformation, only the less strained conformations (whose differential strain energy is lower than a given value, generally 10 kcal mole⁻¹) are refined.

THE FREE ENERGY OF INTERCALATION: THE STRUCTURE OF GRAPHITE INTERCALATION COMPOUNDS

Zhuo-Min Chen, Omar A. Karim,* and B. Montgomery Pettitt

Department of Chemistry, University of Houston, Houston, TX 77204-5641

*Present address: Chemistry Dept., University of North Carolina at Wilmington,
Wilmington, NC

ABSTRACT

A statistical mechanical theory is developed and applied to study the structural effects that the thermodynamic state of alkali ions and small diatomic molecules have on graphite intercalation compounds. The systems considered are that of second stage Rb-graphite and O₂-graphite. Two-dimensional diffraction patterns are computed and compared with experimental measurements. Sensitivity to model parameters is considered. A low order density functional expansion is found to adequately describe the structure of the system modeled as a two-dimensional one-component fluid in an anisotropic external field.

1. INTRODUCTION

Knowledge of the interesting physical and chemical properties of the intercalation compounds of graphite has recently multiplied many-fold due to a renewed experimental interest.¹ Using the techniques of statistical

mechanics,²⁻⁴ theories have been developed and applied to study the effects that ions and molecules have on the structure and thermodynamics of graphite intercalation compounds. These are theories of the free energy or, equivalently, distribution functions for anisotropic liquids and bear a mathematical resemblance to current theories of freezing and crystallization. For the graphite systems, a theoretical understanding of the phenomena of staging (layering) and the concomitant phase transitions from one stage to another involves interlayer distribution functions. As a prelude, the details of the interionic and intermolecular distributions within the graphite layers must first be calculated before the multiple layer problems may be approached at an atomic level of description.

Recent work in this area, from our laboratory, has been devoted to developing and implementing new theoretical methods for exploring the equilibrium properties of graphite intercalation compounds which will be useful in understanding the relation between the structure and the associated physical properties of these compounds, especially the observed catalytic activity.⁵ An accurate structural theory will allow the calculation of the thermodynamics of graphite intercalation compounds and of how free energy differences and ion or molecule solubilities are related to the registration of the graphite layers.

A detailed understanding of these facets of lamellar graphite compounds is of fundamental importance to our understanding of the material properties for which such compounds are useful. These materials are important for a number of energy related systems such as batteries and, in addition, detailed

knowledge of the structure would aid in our understanding of the modulation and regulation of a number of organic reactions which these materials catalyze. It is widely recognized that alkali graphite compounds are active as hydrogenation catalysts for unsaturated hydrocarbons.⁵ These materials are also known to catalyze polymerization in a variety of unsaturated organic compounds.⁶ The most interesting structural aspect of these reactions is that the alkali metal cations are thought to be one of the major centers of the catalytic activity. Thus, large multiatomic reactant molecules must intercalate between the layers of graphite.

Synthetic metals of graphite are made by incorporating alkali atoms between certain layers of a graphite lattice. Such materials display a rich variety of physical properties which are quite distinct from the ordinary or true metals. These compounds have a complicated phase diagram due to the phenomenon of staging.^{1,7} A stage- n compound is defined as having n intervening layers of graphite between the layers of alkali metal atoms. A number of properties vary sharply with a change in stage such as band structure, conductivity (and superconductivity), specific volume and catalytic activity.

An accurate theory of the microscopic structure used with classical statistical thermodynamics would yield a precise picture of the free energy or chemical potential as a function of the geometric and interaction parameters of the system. We shall present ways to formulate and use density functional and integral equation theories of the structure to predict many aspects of these materials' thermodynamic properties based on the coordinate dependent free energy or the distribution functions for the system.

Below, we briefly outline the theory used to find the distribution of intercalant atoms in both atomic and molecular systems. Examples then follow.

II. SIMPLE FLUID THEORY

To make the calculations more tractable, we first use the symmetry of the system to reduce the three-dimensional problem to a two-dimensional one. The model we use for the layers of graphite which act on the intercalated fluid of ions, atoms or molecules is a system composed of a two-dimensional fluid in the presence of an external potential. We have chosen to model the total potential energy for the system, $U(\{\underline{r}_i\})$, as consisting of a sum of one-body and two-body terms. The one-body potential, $U^{(1)}(\underline{r}_i)$, contains the two-dimensional periodically varying interaction of the graphite layers acting on the i^{th} ion. The two-body part of the potential energy of interaction, $U^{(2)}(\underline{r}_i, \underline{r}_j)$, represents the effective interactions between the intercalant species in the system. The total model potential energy of the system is, thereby, taken to be

$$U(\{\underline{r}_i\}) = \sum_i U^{(1)}(\underline{r}_i) + \sum_{i>j} U^{(2)}(\underline{r}_i, \underline{r}_j) \quad 1)$$

where the position of atom (ion) i is denoted \underline{r}_i , and $\{\underline{r}_i\}$ represents the collection of all coordinates \underline{r}_i . In this equation, the summations extend over all ions in the two-dimensional system.

Consistent with experimental observations,^{8,9} we employ a low order Fourier expansion of $U^{(1)}(\underline{r}_i)$ and choose a two-dimensional cartesian coordinate system, (x, y) ,

$$U^{(1)}(\underline{r}) = 2K \left\{ 2 \cos \left[\left(\frac{2\pi}{a} \right) x \right] \cos \left[\left(\frac{2\pi}{a} \right) \frac{1}{\sqrt{3}} y \right] + \cos \left[\left(\frac{2\pi}{a} \right) \frac{2}{\sqrt{3}} y \right] \right\} \quad 2)$$

Here, the lattice constant is given by $a = \sqrt{3}r_e$ where r_e is the equilibrium bond distance between graphitic carbon centers.

The choice of $U^{(2)}$ is governed by the choice of intercalant fluid and by the nature of the electronic interactions between the intercalant and the graphite sheets, giving rise to possible intra-layer screening in the case of ionic species. In the case of alkali atoms, evidence from electronic structure calculations and other sources indicates that there is an electron transfer from the alkali to the graphite.¹²⁻¹⁵ We have taken the two-body interaction between ions to be that of a screened Coulomb form.

$$U^{(2)}(\mathbf{r}_1, \mathbf{r}_2) = q_1 q_2 \exp(-\kappa |\mathbf{r}_1 - \mathbf{r}_2|) |\mathbf{r}_1 - \mathbf{r}_2|^{-1} \quad 3a)$$

As all of the ions are of the same charge, q , this potential is purely repulsive. For realistic choices of the screening parameter, κ , for the ions no hard-core nor Lennard-Jones r^{-12} type of repulsion was found necessary for our qualitative level of comparison with experiment.

We must extend this to the molecular intercalant as well. In principle, our total energy could have more terms which express the chemical bond energy. Alternatively, one might assume that the bond energy could be included into the two-body part of potential. In such a case, the atomic density equations need not be modified at all. However, one would then have to deal with the well-known problem of the different magnitudes of the mean forces responsible for chemical bonding as opposed to physical packing. Alternatively, we assume rigid molecules using an ansatz similar to the RISM theories¹⁰ for molecular fluids and then derive an equation which is consistent with this molecular theory for the singlet distribution of atoms.

For the interactions between molecules a two-site Lennard-Jones potential surface was employed.

$$U^{(2)}(r_i, r_j) = 4\epsilon_{ij} \left(\left[\frac{\sigma_{ij}}{r_{ij}} \right]^{12} - \left[\frac{\sigma_{ij}}{r_{ij}} \right]^6 \right) + \frac{q_i q_j}{r_{ij}} \quad 3b)$$

Here, r_{ij} , ϵ_{ij} , and σ_{ij} , are the distance between atoms i and j , the Lennard-Jones well-depth and diameter, for the atom pair i and j . The diatomic intramolecular potential was taken as a simple rigid bond, which, while convenient, is not crucial to the analysis that follows and could be replaced with a more realistic potential form.

The most graphic data on the structure of graphite intercalation complexes are given by recent diffraction experiments.⁸ Comparison requires only the singlet and pair density distributions. While this sort of information is available from hierarchal equations,¹¹ such an approach would require an approximate closure or truncation of the hierarchy. Instead, the one-body distribution and the pair correlations may be approximated by simultaneously solving the anisotropic Ornstein-Zernike (OZ) equation with an approximate closure expression and an equation that relates the singlet distribution function to the pair distribution function. In full form, this is not completely feasible.

For an anisotropic system we may write the OZ equation as:

$$h^{(2)}(r_1, r_2) = c^{(2)}(r_1, r_2) + \int c^{(2)}(r_1, r_3) \rho(r_3) h^{(2)}(r_3, r_2) dr_3 \quad 4)$$

where $c^{(2)}$ is the OZ direct correlation function and may be taken to be defined by this equation. The pair correlation function is denoted by $h^{(2)}$ and ρ is the

r -dependent singlet density distribution function. A wide variety of approximate closure relations may be used with OZ-like equations. The functional form of the pair potentials chosen make hypernetted chain (HNC)¹² closure a superior choice over alternatives such as the Percus-Yevick (PY) relation.¹¹

The anisotropic HNC closure is given by

$$c^{(2)}(r_1, r_2) = \exp[-\beta U^{(2)}(r_1, r_2) + h^{(2)}(r_1, r_2) - c^{(2)}(r_1, r_2)] \\ - h^{(2)}(r_1, r_2) + c^{(2)}(r_1, r_2) - 1 \quad 5)$$

$\beta = 1/k_B T$ with k_B the Boltzmann constant and T the absolute temperature.

The OZ relation (eq. 4) and the HNC closure (eq. 5) thus give an approximate representation of the pair correlations in an anisotropic fluid given a non-constant $\rho(r)$.

Calculation for a non-isotropic fluid requires a nontrivial (e.g., constant) singlet density distribution and so we may start with the exact Lovett-Mou-Buff equation¹³

$$\nabla_1 \{ \ln[\rho(r_1)] + \beta U^{(1)}(r_1) \} = - \int \rho(r_2) \nabla_2 c^{(2)}(r_1, r_2) dr_2 \quad 6)$$

In principle, eqs. 4, 5 and 6 form a closed set of relations and can be solved for the one- and two-particle correlation functions. Such a solution is quite computer memory intensive and time consuming. In order to circumvent these numerical difficulties, we consider a perturbation expansion of $c^{(2)}(r_1, r_2)$ using an isotropic reference system. We denote quantities relating to the reference system with a subscript '0'. The reference system is an unmodulated liquid with a density ρ_0 identical to the bulk density of the modulated liquid. Hence, $\rho(r)$ must satisfy

$$\frac{1}{V} \int \rho(r) dr = \rho_0 \quad 7)$$

where V is the volume (or area) of integration.

We consider an expansion of the direct correlations about that of the reference system.^{14,15} In terms of the isotropic reference system, the functional Taylor expansion of the direct correlation function is

$$c^{(2)}(r_1, r_2) = c_0^{(2)}(r_{12}) + \int c_0^{(3)}(r_1, r_2, r_3) [\rho(r_3) - \rho_0] dr_3 + \dots \quad 8)$$

To form a solution, we neglect higher order terms involving integrals over $c^{(3)}(r_1, r_2, r_3)$, etc. Such an approximation is not unlike that used in modern density functional theories of freezing.¹⁵⁻¹⁷ In such theories of freezing, the two-particle direct correlation function for the solid is approximated by the corresponding quantity for the coexisting liquid. The difference in using this truncated expansion here versus that in the theories of freezing is that, here, the bulk density and thermodynamic phase of the reference system and the modulated system are necessarily the same. Thus, substituting

$$c^{(2)}(r_1, r_2) \simeq c_0^{(2)}(r_{12})$$

into eq. 6 we obtain

$$\nabla_1 \{ \ln[\rho(r_1)] + \beta U^{(1)}(r_1) \} = - \int \rho(r_2) \nabla_2 c^{(2)}(r_{12}) dr_2 \quad 9)$$

We now invoke the identity

$$\nabla_2 c_0^{(2)}(r_{12}) = - \nabla_1 c_0^{(2)}(r_{12})$$

and substitute into eq. 9 to find, upon interchanging integration and differentiation,

$$\nabla_1 \{ \ln[\rho(r_1)] + \beta U^{(1)}(r_1) \} = \nabla_1 \int \rho(r_2) c_0^{(2)}(r_{12}) dr_2 \quad 10)$$

Equation 10 can be integrated to yield

$$\rho(r_1) = A \exp\{-\beta U^{(1)}(r_1) + \int c_0^{(2)}(r_{12})\rho(r_2)dr_2\} \quad (11)$$

where A is an integration constant to be determined by eq. 7. The similarity between eq. 11 and the corresponding expression for the density functional freezing theories is now evident.

The experimental structure factors are necessarily measured as a function of only one vector argument in momentum space and related to an effective radial distribution function given by

$$G(r) = \frac{1}{\rho_0 N} \int \rho(r_1) g^{(2)}(r_1, r_1 + r) \rho(r_1 + r) dr_1 \quad (12)$$

The experimental structure factor is then related to the Fourier transform of eq. 12,

$$S(k) = 1 + \rho_0 \int e^{-ik \cdot r} (G(r) - 1) dr \quad (13)$$

For our calculation, we perform an expansion of $g^{(2)}$ about an isotropic reference system similar to that in eq. 8 and, again, keep only the first term. This, then, is used in the right hand side of eq. 13 with the nontrivial singlet density for our system obtained from a solution to eq. 11.

III. MOLECULAR FLUID THEORY

To see the path for the molecular case, it is convenient to start from the functional definition of the full two-point correlation function $h(r_1, r_2)$:¹²

$$\frac{\delta \rho(r_1)}{\delta \log z^*(r_2)} = \rho(r_1) \delta(1, 2) + \rho(r_1) \rho(r_2) h(r_1, r_2) \quad (14)$$

where $\log z^*(r)$ may be related to an external activity or potential field, and

$\delta(1,2) = \delta(|r_1 - r_2|)$. For the multicomponent fluid case, we define :

$$\frac{\delta \rho_i(r_1)}{\delta \log z_j^*(r_2)} = \rho_i(r_1) \delta_{ij}(r_1, r_2) + \rho_i(r_1) \rho_j(r_2) h_{ij}(r_1, r_2) \quad (15)$$

where the notational extension is obvious. Here, we introduce the possibility of rigid (covalent) chemical bonding in a neat fluid via $\delta_{ij}(1,2) = \delta(|1-2| - l)$ with the convention that when $i = j, l = 0$; otherwise, if there is a chemical bond, $l = l_o$, the chemical bond length between atoms (or sites) i and j . Since $\rho_i(r)$ is a function of both $\log z_i^*(r)$ and $\log z_j^*(r)$, we have:

$$\begin{aligned} \delta \log \rho_i(1) &= \delta \log z_i^*(1) + \int \delta_{ij}(1,2) \delta \log z_j^*(2) d2 + \\ &\int h_{ii}(1,2) \rho_i(2) \delta \log z_i^*(2) d2 + \int h_{ij}(1,2) \rho_j(2) \delta \log z_j^*(2) d2. \end{aligned} \quad (16)$$

We now let an overline denote a vector and square brackets represent a matrix. Since the Fourier transform and the functional derivative operation commute for well-behaved distributions, we have

$$\overline{\delta \log \rho} = \overline{\delta \log z^*} + [s] * \overline{\delta \log z^*} + [h] * [\rho] \overline{\delta \log z^*} \quad (17)$$

where $[s]$ is the bond or intramolecular correlation function matrix,¹⁰ and $[\rho]$ is a diagonal matrix. Here, $*$ denotes a matrix convolution product. From here we may use the fact that the third term is of higher order in ρ , which is small and can be approximated in a mean field fashion by its average. This leads directly to:

$$\overline{\delta \log \rho} = (I + [s] + \rho_o[h]) * \overline{\delta \log z^*} \quad (18)$$

$$= ([w] + \rho_o[h]) * \overline{\delta \log z^*} \quad (19)$$

where $I + [s] = [w]$ and I is the identity matrix.

From here we use the OZ-like RISM expression¹⁰ to introduce the isotropic molecular site-site pair correlations:

$$[h] = [w] * [c] * [w] + \rho_o [w] * [c] * [h] \quad (20)$$

After an integration by parts of eq. 11 and using eq. 14, the molecular density equation in matrix form is given by:

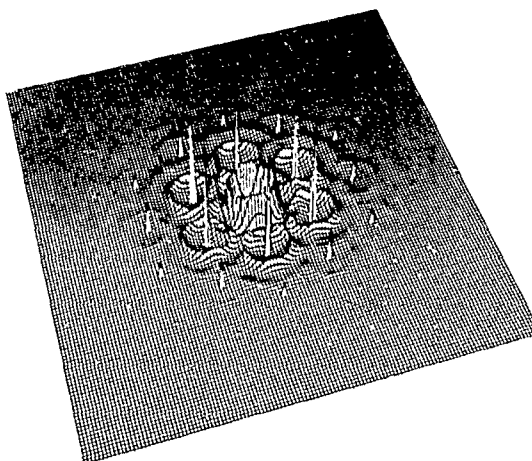
$$\bar{\rho} = A \exp(-\beta \bar{U}_1(r)) + [w] * [c] * \bar{\rho} \quad (21)$$

where A is the normalization constant fixed by the density as in the atomic case.

Again, we may use the HNC equation in matrix element form.¹²

IV. RESULTS

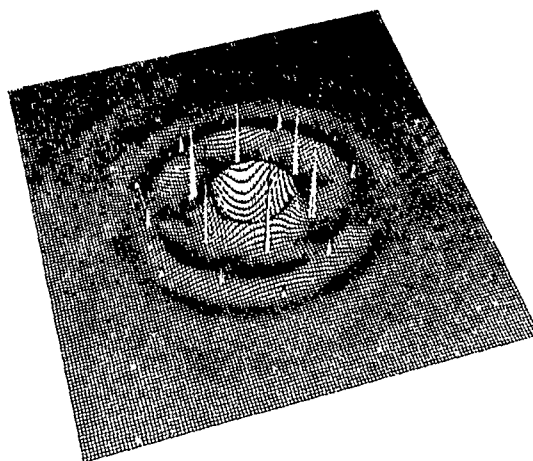
Briefly, we present a comparison of the results for atomic (ionic) intercalant fluids with that for diatomic fluids. The predicted $S(k_x, k_y)$ for the atomic case corresponding to second stage Rb-graphite is shown below.



The periodic density waves in the Rb ions due to the modulation by the graphite are evident in the delta function-like features. Interference of

the simple liquid structure or Debye-Scherer rings with the density waves causes the halos about each of the sharp density wave features. The size and placement of these features is in excellent agreement with those measured experimentally.⁸

In the case of diatomics, there is another length scale in the problem. Besides the size of the intercalant atoms and the underlying graphite lattice there is the length of the bond in the diatomic intercalating fluid. This extra length scale *competes* with the scales defined in the previous picture to change the overall intensity pattern. For physically reasonable bond lengths, we find that this competition has a dramatic effect on the predicted intensity pattern.¹⁸ Below, we show the results for a system corresponding to intercalated O_2 .



The presence of the other length scale is seen in the new features not seen in the previous picture, as well as in the diminution of the features overall. The reduction in intensity is primarily due to destructive interference in the density waves for the fluid atoms of the system.

CONCLUSIONS

The coordinate dependent free energies and the concomitant density distributions for two physically different intercalating fluids have been calculated and compared. We predict that the extra length scales found in some molecular systems will diminish the intensity patterns, even for quite dense fluids. This has been analyzed and found to be due to a competition of the features due to the length scale responsible for chemical bonding with those responsible for packing and the graphite induced density waves.

Our approach has been based on density functional theory using integral equations as input. While such methods are good for structure, dynamics can not be studied in this way. In future work, we shall employ constant energy molecular dynamics and constant chemical potential dynamics to further study these systems structurally as well as dynamically.

Acknowledgements

The authors thank the Robert A. Welch foundation for support of this work. B.M.P. is an Alfred P. Sloan Fellow. Prof. S. C. Moss and Dr. J. D. Fan are acknowledged for many stimulating conversations.

REFERENCES

1. H. Kamimura, *Physics Today*, 40-12, 64(1988).
2. M. Plischke, *Can. J. Phys.*, 59, 802(1981).
3. D. P. DiVincenzo, *Synthetic Metals*, 12, 111(1985).
4. O. A. Karim and B. M. Pettitt, *Chem. Phys. Lett.*, 137, 72(1987); Z.-M. Chen, O. A. Karim and B. M. Pettitt, *J. Chem. Phys.*, 89, 1042 (1988).
5. K. Tamaru, *Catal. Rev.*, 4, 161(1970).
6. I. M. Panayotou and I. B. Rashkou, *J. Poly. Sci.*, 10A-I, 1267(1972).

7. T. Ohno and H. Kamimura, *J. Phys. Soc. Japan*, 52, 223(1983).
8. F. Rousseaux, R. Moret, D. Guerard, P. Lagrange and M. Lelaurin, *Synth. Met.*, 12, 45(1985).
9. S. C. Moss, G. Reiter, J. L. Robertson, C. Thompson, J. D. Fan and K. Ohskima, *Phys. Rev. Lett.*, 57, 3191(1986).
10. D. Chandler and H.C. Andersen, *J. Chem. Phys.*, 57, 1930 (1972); D Chandler, in "The Liquid State of Matter: Fluids Simple and Complex", ed. E.W. Montroll and J.L. Lebowitz, (North Holland, Amsterdam 1982) pg. 275.
11. J. P. Hansen and I. R. McDonald, "Theory of Simple Liquids", (Academic Press, New York, 1976).
12. J.M.J. Van Leeuwen and J. Groeneveld, J. DeBoer, *Physica*, 25, 792 (1959); J. DeBoer, J.M.J. Van Leeuwen and J. Groeneveld, *Physica*, 30, 2265 (1964).
13. R. Lovett, C. Y. Mou and F. P. Buff, *J. Chem. Phys.*, 65, 570(1976).
14. S. Sokolowski and W. Steele, *J. Chem. Phys.*, 82, 2499(1985).
15. T. V. Ramakrishnan and M. Yussouff, *Phys. Rev. B*, 14, 2775(1979).
16. D. W. Oxtoby and A. D. J. Haymet, *J. Chem. Phys.*, 74, 2559(1981).
17. K. Ding, D. Chandler, S. J. Smithline, A. D. J. Haymet, *Phys. Rev. Lett.*, 59, 1698(1987); S. J. Smithline, S. W. Rick and A. D. J. Haymet, *J. Chem. Phys.*, 88, 2004(1988).
18. Z.-M. Chen and B.M. Pettitt, *J. Chem. Phys.* (submitted).

DISCUSSION

SQUMPASIS - Do you assume that the 2-D liquid 'sheets' are completely decoupled through the graphite layers ?

PETTITT - No, but that information has been integrated out of the picture. Thus, it appears only in the one-body potential we use. The potential was, however, obtained for 2nd stage Rb⁺ - graphite (which has a specific coupling or staging of the layers) from the intensity of the Bragg spots induced by the liquid density waves. While averaged effects are there, we cannot ask specific questions about correlations between layers with our current model.

BUCKINGHAM - Is there a threshold in the concentration of alkali atoms in the graphite intercalation materials ? If the layers of carbon atoms remain approximately parallel, then it would seem that there will be a threshold concentration of intercalate.

PETTITT - Yes, there are in fact several thresholds. These are, again, related to this phenomena of staging. There is a phase transition from one stage to another and each new stage brings with it a new set of physical properties such as color, conductivity, etc.

MONQUE - Because in your model you talk about the "tailored layer expansion factor" and about the way you can calculate it, you've made me think on its application for the synthesis or preparation of pillared clays (catalytic material built with some different cations in interlayer position with specific shape selectivity as a function of the interlayer space). So can you please tell me how can I use this model experimentally on the synthesis of this type of materials ?

PETTITT - Pillared clays and some zeolites are certainly good examples of solid supports with liquid accessibility for catalysis. Our understanding of even these simple graphite intercalation systems is not yet sufficient to make thermodynamic suggestions. Hopefully, we shall someday understand the layer spacing free energy well enough to make such suggestions.

CHEMISORPTION ON TRANSITION METAL SURFACES AND HETEROGENEOUS CATALYSIS: BACK-OF-THE-ENVELOPE THEORETICAL MODELING

EVGENY SHUSTOROVICH

Corporate Research Laboratories, Eastman Kodak Company, Rochester, New York
14650-2001, U.S.A.

SUMMARY

The lecture highlights the efficiency of the "back-of-the-envelope" BOC-MP (bond-order conservation Morse-potential) model in calculating reaction energetics on metal surfaces, particularly heats of adsorbate chemisorption Q and activation barriers ΔE^* for adsorbate dissociation and recombination. After representative calculations of Q and ΔE^* for a variety of diatomic and polyatomic ad molecules, we discuss an example of a complex heterogeneous catalytic process — transformations of C_2 hydrocarbons on transition-metal surfaces. It appears that the BOC-MP modeling provides an effective means for understanding and projecting surface reactivity.

INTRODUCTION

The worldwide value of products made via catalytic technology is in excess of one trillion dollars per year, and heterogeneous catalysis leads the way. This tremendous economic impact is the major thrust of intensive efforts in both industry and academia to understand heterogeneous catalytic processes at the molecular level. Understanding mechanisms of chemical reactions is impossible without knowledge of reaction energetics. Ultimately, this energetics should be calculated by quantum-mechanical techniques. So far, however, the progress has been impressive for gas-phase reactions but rather modest for surface reactions.

While waiting for the advent of efficient quantum-mechanical techniques, common sense requires one to look for working alternatives. If not microscopic modeling, let it be phenomenological. To have a chance to succeed, such a phenomenological model should have a rigorous mathematical framework and make use of well-defined parameters, preferably only observable ones. In order to be practical, theoretical constructs should be simple enough, ideally at a "back-of-the-envelope" level.

During the last years, we have been developing such a phenomenological construct, namely, the bond-order conservation Morse-potential (BOC-MP) model (refs. 1-3). The model is based on four assumptions:

(1) Each two-center metal-adsorbate $M-A$ interaction is described by the Morse potential!

$$E(x) = -Q(x) = -Q_0(2x - x^2) \quad (1)$$

where x is the M-A bond order,

$$x = \exp[-(r - r_0)/a] \quad (2)$$

an exponential function of the M-A distance r (r_0 and a are constants), and Q_0 is the M-A equilibrium bond energy. Obviously, the total energy $E(x)$ (Eq. 1) has only one minimum at the equilibrium distance r_0 when the bond order $x = 1$, by definition (Eq. 2).

(2) For an adatom A interacting with n metal atoms M (M_n -A), n two-center M-A interactions are additive.

(3) For a given M_n -A, n is limited to nearest neighbors.

(4) For $X = A$ or AB, along a migration path up to dissociation, the total M_n -X bond order is conserved and normalized to unity.

The assumptions (1) - (4) are the rules of the game. The rest is straightforward algebra of maximizing the total M_n -X bond energy under the BOC condition (4) the analytic form of which depends on the nature of X and the geometry of M_n .

The aim of the present lecture is to give a flavor of how the BOC-MP model treats the energetics of diatomic and polyatomic adsorbates on transition-metal surfaces, particularly the heat of chemisorption Q and activation barriers ΔE^* for dissociation and recombination. Because of the space constraints, the analytic formulas used to calculate Q and ΔE^* will be given with minimal explanations, and only a few sources of the experimental data employed will be cited. For a detailed discussion and a complete list of references, the reader is referred to the review (ref. 3).

THE BOC-MP FORMALISM

Heats of Chemisorption

For atomic chemisorption, the M_n -A bond energy Q_n monotonically increases with n as

$$Q_A = Q_n = Q_{0A}(2 - 1/n) \quad (3)$$

where Q_{0A} is the maximum M-A two-center bond energy (cf. Eq. 1). The value of Q_n reaches the absolute maximum in the hollow n -fold site, so that the observed heat of atomic chemisorption Q_A can be identified with Q_n .

For molecular chemisorption, one should distinguish between weak and strong M_n -AB bonding. The weakly bound AB molecules typically have a closed electronic shell, e.g., H_2 , N_2 , CO , NH_3 , H_2O , or unpaired electrons occupying the substantially delocalized molecular orbitals, e.g., NO or O_2 . The strongly bound AB molecules have unpaired electrons retaining their atomic character, e.g., CH , CH_2 , NH , OH , or OCH_3 .

Analytically, the difference between the weak and strong M_n -AB bonding is reflected in the use of different effective Morse constants, Q_{0A} and Q_A , respectively.

For the weak M_n -AB bonding, the simplest case corresponds to AB perpendicular to a surface with the A end down when, to first approximation, the M_n -B bond order can be neglected. For such mono(η^1) coordination via A with M_n ($\eta^1\mu_n$), the variational procedure gives

$$Q_{AB,n} \leq \frac{Q_{0A}^2}{(Q_{0A}/n) + D_{AB}} \quad \text{for } D_{AB} > \frac{n-1}{n} Q_{0A} \quad (4)$$

If AB is coordinated via both A and B (dicoordination η^2), the bonding energy becomes

$$Q_{AB} = \frac{ab(a+b) + D_{AB}(a-b)^2}{ab + D_{AB}(a+b)} \quad (5)$$

where

$$a = Q_{0A}^2(Q_{0A} + 2Q_{0B}) / (Q_{0A} + Q_{0B})^2$$

and

$$b = Q_{0B}^2(Q_{0B} + 2Q_{0A}) / (Q_{0A} + Q_{0B})^2$$

For a homonuclear A_2 ($a = b = (3/4)Q_{0A}$), Eq. 5 reduces to

$$Q_{A_2} = \frac{(9/2)Q_{0A}^2}{3Q_{0A} + 8D_{A_2}} \quad (6)$$

For the strong M_n -AB bonding in the monocoordination $\eta^1\mu_n$, the variational procedure now leads to an expression

$$Q_{AB} = \frac{Q_A^2}{Q_A + D_{AB}} \quad (7)$$

which is an analog of Eq. 4 (for $n = 1$) where Q_{0A} is substituted by Q_A . Along with the weak and strong M_n -AB bonding, one can imagine the intermediate one, which may be described by interpolating between the two extremes. In particular, for the monocoordination ($\eta^1\mu_n$) M_n -AB, one can simply average Eqs. 4 and 7 as

$$Q_{AB} = 1/2 \left[\frac{Q_{0A}^2}{(Q_{0A}/n) + D_{AB}} + \frac{Q_A^2}{Q_A + D_{AB}} \right] \quad (8)$$

The intermediate M_n -AB bond strength is expected for monovalent radicals AB where A is a tri- or tetravalent atom, say N or C in NH_2 , CH_3 , HCO .

Dissociation and Recombination Barriers

If AB approaches a surface from the gas phase, the multidimensional activation barrier $\Delta E_{AB,g}^*$ for dissociation $AB_g \rightarrow A_s + B_s$ explicitly depends on the chemisorption energies of all adsorbates, namely, the barrier can be approximated as

$$\Delta E_{AB,g}^* = 1/2(\Delta E_{AB,g}^{**} - Q_{AB}) \quad (9)$$

where the one-dimensional barrier $\Delta E_{AB,g}^{**}$, rigorously defined by variational procedure, is

$$\Delta E_{AB,g}^{**} = D_{AB} - (Q_A + Q_B) + \frac{Q_A Q_B}{Q_A + Q_B} \quad (10)$$

Obviously, the dissociation barrier $\Delta E_{AB,s}^*$ from a chemisorbed state will be larger than $\Delta E_{AB,g}^*$ (one- or multidimensional) just by the amount of the molecular heat of chemisorption Q_{AB} :

$$\Delta E_{AB,s}^* = \Delta E_{AB,g}^* + Q_{AB} \quad (11)$$

Combining with Eq. 9, we obtain

$$\Delta E_{AB,s}^* = 1/2(\Delta E_{AB,g}^{**} + Q_{AB}) \quad (12)$$

In general, for the reverse reaction of recombination, the activation barriers can be calculated from the relevant thermodynamic relations. Specifically, for the recombination of chemisorbed A_s and B_s to chemisorbed AB_s or gas-phase AB_g , the activation barriers $\Delta E_{A-B,s}^*$ and $\Delta E_{A-B,g}^*$ may be the same or different depending on the sign of the gas-phase dissociation barrier $\Delta E_{AB,g}^*$, namely,

$$\Delta E_{A-B,s}^* = \Delta E_{A-B,g}^* = Q_A + Q_B - D_{AB} + \Delta E_{AB,g}^* \quad \text{if } \Delta E_{AB,g}^* > 0 \quad (13)$$

or

$$\Delta E_{A-B,g}^* = \Delta E_{A-B,s}^* - \Delta E_{AB,g}^* = Q_A + Q_B - D_{AB} \quad \text{if } \Delta E_{AB,g}^* < 0 \quad (14)$$

For some reactions, for example, $\text{CO}_s + \text{O}_s \rightleftharpoons \text{CO}_{2,s}$ or $\text{NO}_s + \text{N}_s \rightarrow \text{N}_2\text{O}_s$, the one-dimensional treatment is satisfactory (ref. 3). In this case, the recombination barrier is particularly simple (cf. Eqs. 10 and 13):

$$\Delta E_{A-B}'' = \frac{Q_A Q_B}{Q_A + Q_B} \quad (15)$$

Here, if $Q_A \gg Q_B$, $\Delta E_{A-B}''$ will be close to Q_B , the heat of chemisorption of the weaker bound partner.

In conclusion, two general points should be stressed. First, in Eqs. (4)–(15), A and B may be either atoms or atomic groups, so that the same formalism is used to calculate Q and ΔE^* for both diatomic and polyatomic molecules (at the zero-coverage extreme). Second, the BOC-MP analytic interrelations make use of observables only, namely, the heats of atomic chemisorption (Q_A, Q_B) and various constants—thermodynamic (D_{AB}), structural (n), and numerical (coefficients). For these reasons, the BOC-MP model can treat a remarkably broad variety of ad molecules, and comparison of the model projections with experiment is particularly straightforward. Consider now some representative examples.

BOC-MP Applications

Table 1 lists heats of chemisorption of various ad molecules — diatomic and polyatomic, mono- and dicoordinated, weakly and strongly bound. We see that agreement with experiment is excellent, typically within 1–2 kcal/mole.

Table 2 summarizes the activation barriers of some well-studied reactions of dissociation and recombination, in good agreement with experiment. In particular, for the recombination reactions $\text{CO}_s + \text{O}_s \rightarrow \text{CO}_{2,s}$ and $\text{NO}_s + \text{N}_s \rightarrow \text{N}_2\text{O}_s$, as Eq. 15 predicts, the activation barriers follow closely the values of Q_{CO} and Q_{NO} , respectively, but not Q_{O} and Q_{N} .

TABLE 1

Initial Heats of Molecular Chemisorption Q_{AB} ^a

Surface	Coord. Type	AB	Experimental Values of			Q_{AB}	
			Q_A	Q_B	D_{AB} ^b	Calcd.	Exp.
Ni(111)	η^1	CO	171		257	29 ^c	27
Pt(111)	η^1	NO	116		150	26 ^d	27
Pd(111)	η^1	NO	130		150	32 ^d	31
Pt(111)	η^1	NH ₃	116		279	13 ^c	12-15
Ni(111)	η^1	NH ₃	135		279	18 ^c	20
Pt(111)	η^1	OH ₂	85		220	11 ^c	12
Pt(111)	η^1	O=CH ₂	85		176	11 ^c	11
Pt(111)	η^2	O ₂	85	85	119	11 ^e	9
Ru(001)	η^2	O=C(CH ₃) ₂	100	67	179	15 ^f	16
Ni(111)	η^2	H ₂ C=CH ₂	171	171	355	14 ^e	13
Pt(111)	η^1	OH	85		102	399	36-45
Pt(111)	η^1	NH	116		102	689	63-69

^aSee Ref. 3 for sources of the experimental values of Q . All energies in kcal/mole.^bRef. 4.^cEquation 4 for $n = 1$.^dEquation 4 for $n = 2$.^eEquation 6.^fEquation 5.^gEquation 7.

TABLE 2

Dissociation and Recombination Barriers ΔE^* for Some Surface Reactions^a

Reaction	Surface	Experimental Values of				ΔE^*	
		D_{AB}^b	Q_A	Q_B	Q_{AB}	Calcd.	Exp.
$H_{2,g} \rightarrow H_s + H_s$	Fe(111)	104	62	62	7	2 ^c	0
	Ni(111)		63	63	7	1 ^c	2
	Cu(100)		56	56	5	7 ^c	5
$N_{2,g} \rightarrow N_s + N_s$	Fe(110)	228	138	138	8	6 ^c	8
	Fe(100)		140	140	8	4 ^c	2
	Fe(111)		139	139	8	5 ^c	0
$CO_g \rightarrow C_s + O_s$	Ni(111)	257	171	115	27	6 ^c	
	Ni(100)		171	130	30	0 ^c	-3
	W(110)		200	125	21	-6 ^c	-15
$NO_g \rightarrow N_s + O_s$	Rh(100)	151	128	102	26	-23 ^c	-15
	Pt(111)		116	85	27	-14 ^c	
$CO_{2,g} \rightarrow CO_s + O_s$	Rh(111)	127	32	102		17 ^d	17
	Re(001)		29	127		-5 ^d	≤ 0
$CO_s + O_s \rightarrow CO_{2,g}$	Rh(111)		32	102		24 ^d	27
	Pd(111)		34	87		24 ^d	25
	Pt(111)		32	85		23 ^d	25
	Ag(110)		6.5	80		6.0 ^d	5.3
$NO_s + N_s \rightarrow N_2O_s^e$	Rh(111)		26	128		22 ^d	21
	Rh(100)		25	131	21	21 ^d	21
	Pt(111)		27	116		22 ^d	20

^aSee Ref. 3 for sources of the experimental values of Q and ΔE^* . All energies in kcal/mole.^bRef. 4.^cEquation 9.^dEquations 10 or 15, respectively.^eFollowed by nonactivated decomposition $N_2O_s \rightarrow N_{2,g} + O_s$.

As an example of a complex heterogeneous catalytic process including several competing pathways, we choose surface reactions of hydrocarbons C_2H_x , which have drawn a great deal of interest because of their fundamental and practical importance.

Naturally for chemists, we are mostly interested in periodic (relative) changes in catalytic behavior of transition metals. Specifically, we select the sequence from Pt to Ni and further to Fe or W, the latter two being simulated as the same model metal Fe/W (with the model parameters averaged over the (close) parameters of real Fe and W). The BOC-MP calculations have been made for the smoothest (most densely packed) surfaces, namely for fcc Pt(111), Ni(111), and bcc Fe/W(110).

Here there are two problems of special interest: (1) the thermochemistry of C_2H_x species in the chemisorbed vs gas-phase state, and (2) the effects of metal composition and the structure of C_2H_x species on the activation energy for C-H and C-C bond cleavage.

Table 3 lists total bond energies in the gas phase (D) and chemisorbed (D + Q) states for all C_2H_x species ($x = 0-6$). Some comparisons of enthalpies in the gas phase vs chemisorbed states are made in Table 4. The calculated activation barriers ΔE^* for C-C and C-H bond cleavage and recombination for chemisorbed C_2H_x species are summarized in Table 5. All the discussion below will refer to *chemisorbed* species if not stated otherwise.

Of general model conclusions, the most important is the following: many reorganizations of C_2H_x species, being highly endothermic in the gas phase, typically become exothermic on transition-metal surfaces. For example, in the gas phase the ground state of C_2H_3 is vinyl $H_2C=CH$ which is lower than ethynyl H_3C-C (the excited state) by 45 kcal/mole. Under chemisorption, this isomerization is highly exothermic ($\Delta H = -8, -15, -25$ kcal/mole for Pt, Ni, Fe/W, respectively). Similarly, isomerization of acetylene to vinylidene $HC\equiv CH \rightarrow H_2C=C$ is highly endothermic in the gas phase ($\Delta H = 44$ kcal/mole) but becomes highly exothermic by 13-41 kcal/mole on the metal surfaces studied. This makes it comprehensible why H_3CC and H_2CC are often observed in the chemisorbed states (unlike the gas phase).

TABLE 3

Total Bond Energies in the Gas-Phase (D) and Chemisorbed (D + Q) States of C_2H_x on Pt(111), Ni(111), and Fe/W(110)^a

C_2H_x	$D_{C_2H_x}$ ^b	$Q_{C_2H_x}$ ^c			$D_{C_2H_x} + Q_{C_2H_x} + (8-x)Q_H$ ^d		
		Fe/W	Ni	Pt	Fe/W	Ni	Pt
H ₃ C-CH ₃	674	6	5	5	812	805	801
H ₃ C-CH ₂	576	64	49	39	838	814	798
H ₃ C-CH	466	107	85	70	837	803	780
H ₂ C=CH ₂	538	20	15	12	822	805	794
H ₃ C-C	376	141	115	97	847	806	778
H ₂ C=CH	421	71	55	44	822	791	770
H ₂ C=C	348	110	87	71	854	813	785
HC≡CH	392	25	18	14	813	788	772
HC≡C	259	106	84	69	827	784	755
CH ₃ + CH ₃	586	124	96	76	842	808	784
CH ₃ + CH ₂	476	166	131	106	840	796	765
CH ₃ + CH	374	204	164	135	842	790	753
CH ₃ + C	293	262	219	188	885	827	786
CH ₂ + CH ₂	366	208	166	136	838	784	746
CH ₂ + CH	264	246	199	165	840	778	734
CH ₂ + C	183	304	254	218	883	815	767
CH + CH	162	284	232	194	842	772	722
CH + C	81	342	287	247	885	809	755
C + C	0	400	342	300	928	846	788
CH ₄ + CH ₄	796	12	12	12	808	808	808

^a The parameters used: $Q_C = 150, 171, 200$ and $Q_H = 61, 63, 66$ for Pt, Ni, Fe/W, respectively (ref 3). All energies in kcal/mole.

^b From Ref. 4 with corrections and additions specified in ref. 3.

^c Equations 4, 7 or 8.

^d Normalized for stoichiometry C_2H_8 ($C_2H_6 + H_2 \rightarrow 2CH_4$), when for C_2H_x (or $CH_y + CH_{x-y}$) the rest $(8-x)$ atoms H are assumed to be atomically chemisorbed.

TABLE 4

Enthalpies of C_2H_x Isomerization Reactions: Gas-Phase vs Chemisorbed States

		ΔH , kcal/mole ^a			
Reaction		Gas Phase	Fe/W(110)	Ni(111)	Pt(111)
C ₂ H ₄	H ₂ C=CH ₂ → H ₃ C=CH	72	-15	2	14
C ₂ H ₃	H ₂ C=CH → H ₃ C-C	45	-25	-15	-8
C ₂ H ₂	HC≡CH → H ₂ C=C	44	-41	-25	-13

^a From Table 3.

TABLE 5

Activation Barriers for Forward ΔE_f^* and Reversed ΔE_r^* Reactions of Chemisorbed C_2H_x ^a

C_2H_x	Reaction	D_{CX} ^b	ΔE_f^{*c}			ΔE_r^{*d}		
			Fe/W	Ni	Pt	Fe/W	Ni	Pt
C_2H_6	$CH_3CH_{3,g} \rightleftharpoons CH_3CH_2 + H$	98	-3	5	8	38	19	10
	$CH_3CH_3 \rightleftharpoons CH_3CH_2 + H$	98	3	10	13	35	19	10
C_2H_5	$CH_3CH_2 \rightleftharpoons CH_3 + CH_2$	100	18	24	33	20	6	0
	$\rightleftharpoons CH_2CH_2 + H$	38	16	11	4	0	2	0
	$\rightleftharpoons CH_3CH + H$	110	21	24	25	20	13	7
C_2H_4	$CH_2CH_{2,g} \rightleftharpoons CH_2 + CH_2$	172	-2	17	36	36	11	0
	$\rightleftharpoons CH_2CH + H$	117	-3	7	13	20	8	1
	$CH_2CH_2 \rightleftharpoons CH_2 + CH_2$	172	18	32	48	34	11	0
	$\rightleftharpoons CH_2CH + H$	117	17	22	25	17	8	1
	$CH_3CH \rightleftharpoons CH_3 + CH$	92	19	23	27	24	10	0
	$\rightleftharpoons CH_2CH + H$	45	25	21	18	10	9	8
	$\rightleftharpoons CH_3C + H$	90	17	19	20	27	22	18
C_2H_3	$CH_2CH \rightleftharpoons CH_2 + CH$	157	21	31	38	39	18	2
	$\rightleftharpoons CHCH + H$	29	14	9	5	5	6	7
	$\rightleftharpoons CH_2C + H$	73	5	7	9	37	29	24
	$CH_3C \rightleftharpoons CH_3 + C$	83	5	8	11	43	29	19
	$\rightleftharpoons CH_2C + H$	28	2	15	13	9	22	20
C_2H_2	$CHCH_g \rightleftharpoons CH + CH$	230	-4	19	36	54	21	0
	$\rightleftharpoons CHC + H$	133	-12	1	11	39	18	8
	$CHCH \rightleftharpoons CH + CH$	230	21	37	50	50	21	0
	$\rightleftharpoons CHC + H$	133	13	19	25	27	18	8
	$CH_2C \rightleftharpoons CH_2 + C$	165	20	27	32	49	29	14
	$\rightleftharpoons CHC + H$	89	34	32	31	4	3	1
C_2H	$CHC \rightleftharpoons CH + C$	178	13	22	29	71	47	29

^a The barriers for the gas-phase ethane, ethylene, and acetylene are also added. All energies in kcal/mole.

^b The difference between the gas-phase total bond energies of the reactant and products (see Table 3).

^c Equations 9 or 12. The values of Q from Table 3.

^d Equations 13 or 14.

For ethylene C_2H_4 on Pt(111) and Ni(111), our estimates for the heats of chemisorption are $Q_{C_2H_4} = 12$ and 15 kcal/mole, respectively, in excellent agreement with the experimental range $Q_{C_2H_4} = 11 - 13$ kcal/mole for Pt, Pd, Ru, and Ni. Consistently, for the metal range Ni-Pt, we find the distinct gas-phase C-H bond dissociation barrier $\Delta E_{CH,g}^* = 7-13$ kcal/mole, which explains why the chemisorbed molecular C_2H_4 requires elevated temperatures for chemical transformations. As seen from Table 3, the formation of CH_3C from C_2H_4 is moderately endothermic on Pt, practically thermoneutral on Ni, and highly exothermic on Fe/W ($\Delta H = 16, -1, -25$ kcal/mole, respectively). On Pt(111), we predict CH_3C to be rather stable since the calculated C-C bond scission barrier is $\Delta E_{CC}^* = 11$ kcal/mole. But on Ni(111) and Fe/W(110), we predict this barrier to be smaller — 8 and 5 kcal/mole, respectively.

Thus the model conclusion is that the stability of CH_3C decreases in the order Pt > Ni > Fe/W. Indeed, ethynylidyne has been observed readily under decomposition of ethylene on Pt, Pd, Rh, Ru (ref. 5), only at high adsorbate coverages on Ni (ref. 6) and, probably on Co (ref. 7), but not on more active metals. One can add that on Pt(111), the C-C bond cleavage $CH_3C \rightarrow CH_3 + C$ ($\Delta E_{CC}^* = 11$ kcal/mole) seems to be more favorable than the C-H bond cleavage $CH_3C \rightarrow CH_2C + H$ ($\Delta E_{CH}^* = 13$ kcal/mole), so that the molecule will retain most of its hydrogen up to the point of C-C bond scission, in agreement with the ^{13}C NMR experiment (ref. 8).

We predict that acetylene chemisorbs slightly stronger than ethylene with $Q_{C_2H_2} = 14, 18, 25$ kcal/mole for Pt(111), Ni(111), Fe/W(110), respectively. Experimental data on $Q_{C_2H_2}$ (usually from TPD spectra) are not available because C_2H_2 begins to decompose before it desorbs. Consistently, we found the gas-phase C-H bond cleavage to be nonactivated on Ni and especially on Fe/W, namely $\Delta E_{CH,g}^* = 1$ and -12 kcal/mole, respectively. On Pt, where $\Delta E_{CH,g}^* = 11$ kcal/mole, the first surface reaction appears to be the distinctly exothermic isomerization $CHCH \rightarrow CH_2C$ for which we found $\Delta H = -13$ kcal/mole. Indeed, the formation of the vinylidene CH_2C intermediate on Pt(111) was first suggested from EELS spectra and confirmed on Pt particles by the ^{13}C NMR analysis (ref. 9). However, this isomerization is not a favorable route for C-C bond scission since $CH_2C \rightarrow CH_2 + C$ would require $\Delta E_{CC}^* = 32$ kcal/mole. So, we predict that dehydrogenation $CHCH \rightarrow CHC + H$ will occur instead ($\Delta E_{CH}^* = 25$ kcal/mole), and then the C-C bond scission $CHC \rightarrow CH + C$ ($\Delta E_{CC}^* = 29$ kcal/mole). Thus, contrary to decomposition of CH_3C , one can expect a substantial loss of hydrogen before the C-C bond rupture, again in agreement with the ^{13}C NMR data (ref. 9).

On Fe/W(110), we project the rather different situation because for $CHCH$ not only the C-H but also C-C bond cleavage seems to be nonactivated from the gas

phase ($\Delta E_{CC,g}^* = -4$ kcal/mole). Thus, on Fe surfaces, one can expect acetylene to rapidly decompose into CH_x fragments. This model projection is consistent with the fact that under heating of chemisorbed CHCH, only CH_x intermediates have been observed on various Fe surfaces (ref. 10). Between Fe/W and Pt, the decomposition products may be a variety of C_2H_x and CH_x species depending on metal composition and reaction conditions. For example, on Ni surfaces, rapid decomposition of CHCH to (partly) HCC and (mainly) CH_x species was reported (ref. 11). At the same time, on Ru(001), whose activity is intermediate between Pt(111) and Ni(111), the whole set of H_xCC species, $x = 1, 2, 3$, resulting from isomerization ($CHCH \rightarrow CH_2C$), dehydrogenation ($CHCH \rightarrow CHC + H$) and rehydrogenation ($CHCH + H \rightarrow [CH_2CH] \rightarrow CH_3C$), was identified by EELS (ref. 12).

The model can also shed light on regularities of hydrogenolysis $C_2H_6 + H_2 \rightarrow 2CH_4$. In general, by comparing possible C-C bond scission routes on Pt vs Ni vs Fe/W (cf. Table 5), one can easily see that the hydrogen content (x) in the hydrocarbon species C_2H_x undergoing this scission decreases in the order Fe/W > Ni > Pt, in agreement with vast experimental observations (ref. 13).

The similar analyses have been successfully done for other heterogeneous catalytic processes such as CO hydrogenation to methane CH_4 and methanol CH_3OH , and decomposition of formic acid $HCOOH$ (ref. 3).

CONCLUSION

The analytic BOC-MP model effectively treats reaction energetics on transition-metal surfaces. This efficiency is of particular importance because the model is based on a few assumptions, within which the model interrelations are exact for atomic adsorbates and well defined for molecular adsorbates, the same analytic formalism being used to treat both diatomic and polyatomic molecules. Moreover, these interrelations are expressed in terms of observables only (the heats of chemisorption and various constants), which makes comparison with experiment typically direct and unambiguous. The rigor and simplicity of this back-of-the-envelope model makes it a promising tool for exploring surface reactivity.

REFERENCES

1. E. Shustorovich, Chemisorption phenomena: analytic modeling based on perturbation theory and bond-order conservation, *Surf. Sci. Rep.*, 6(1) (1986) 1-63.
2. E. Shustorovich, Chemisorption theory: in search of the elephant, *Acc. Chem. Res.*, 21 (1988) 183-189.

3. E. Shustorovich, The bond-order conservation approach to chemisorption and heterogeneous catalysis: applications and implications, *Adv. Catal.*, 37 (1989), in press.
4. CRC Handbook of Chemistry and Physics, Boca Raton, Florida: CRC Press, 1984-1985, pp. F171-190.
5. See a discussion in: T. B. Beebe, Jr. and J. T. Yates, Jr., Infrared spectroscopic investigation of the ethylene chemisorption reaction on supported metallic catalyst surfaces: ethylidyne formation on Pt, Rh, Pd, and Ru supported on alumina, *J. Phys. Chem.*, 91 (1987) 254-257.
6. X.-Y. Zhu and J. M. White, Interaction of ethylene and acetylene with Ni(111): a SSIMS study, *Surf. Sci.*, 214 (1989) 240-256.
7. K. G. Anderson and J. G. Ekerdt, Hydrocarbon surface fragments over Co/SiO₂: an FTIR study, *J. Catal.*, 116 (1989) 556-567.
8. P.-K. Wang, C. P. Slichter, and J. H. Sinfelt, Structures and reactions of C₂H₄ adsorbed on small Pt clusters, *J. Phys. Chem.*, 89 (1985) 3606-3609.
9. P.-K. Wang, C. P. Slichter, and J. H. Sinfelt, NMR Study of the structure of simple molecules adsorbed on metal surfaces: C₂H₂ on Pt, *Phys. Rev. Lett.*, 53(1) (1984) 82-85.
10. See a discussion in: U. Seip, M.-C. Tsai, J. Küppers and G. Ertl, Interaction of acetylene and ethylene with an Fe(111) surface, *Surf. Sci.*, 147 (1984) 65-88.
11. See, for example: J. A. Stroscio, S. R. Bare, and W. Ho, The chemisorption and decomposition of ethylene and acetylene on Ni(110), *Surf. Sci.*, 148 (1984) 499-525.
12. See a discussion in: P. Jakob, A. Cassuto, and D. Menzel, Adsorption and reaction of acetylene on Ru(001): vibrational spectroscopy and thermal desorption, *Surf. Sci.*, 187 (1987) 407-433.
13. J. H. Sinfelt, Some recent developments in catalysis by metals, *J. Phys. Chem.*, 90 (1986) 4711-4723.

DISCUSSION

DURUP - 1. It looks paradoxical that the effect of the environment on the metal atom (crystallographic surface, etc.) directly enters your theory through the atomic adsorption energy, whereas the adsorbed atom environment enters the metal bond energy only indirectly, although it is a major perturbation.

2. Since you called on Karl Popper, could you give an example of a clear-cut prediction in a case where the experiment has not yet been performed?

SHUSTOROVICH -

1. I don't see any paradox. The environment of the adsorbed atom A in the coordinated molecule AB enters via the gas-phase dissociation energy D_{AB} , similar to that of the metal atom M via the atomic chemisorption energy Q_A . Actually an interplay of these energetic parameters, Q_A (and/or Q_B) and D_{AB} determines the heat of molecular chemisorption Q_{AB} [cf. eqs. (4)-(8)].

2. The Popperian criterion or falsifiability requires to make model predictions that can be proved to be wrong. In a sense, all the calculated values of Q_{AB} and ΔE^*_{AB} given above are model predictions since they are obtained from the same set of equations with the fixed parameters. As an example of the clear-cut prediction awaiting the experimental verdict one can take the thermochemistry of C_2H_x transformations (see Tables 3 and 4).

BONIFASSI -

1. Are your results on chemisorption heat different from those obtained by R. Hoffmann for CO on Ni(111) (-1.66 eV) by tight binding methods [J.A.C.S., 1985, 107, 578-584]?

2. Can you precise the principle of bond order conservation?

SHUSTOROVICH -

1. Since the BOC-MP model is different, the more informative question would be how close the calculated values of Q_{CO} are to the experimental value of ~1.2 eV. Our value is 1.3 eV (cf Table 1).

2. If you mean the justification of the BOC principle, it can only refer you to thorough studies, both computational (ab initio) and experimental (X-ray), of a variety of linear three-center systems A-B-C, where the additive BOC at unity ($X_{AB} + X_{BC} = 1$) has invariably been found (see a list of references in the review (ref. 3)). The BOC

condition used in our work is an extension of this well-established regularity to polycenter interactions of quasi-spherical character (where the energy depends only on the distance).

ANGYAN - 1. What was the bond order definition of the ab-initio calculations you mentioned which justify the bond order conservation ?

2. I should like to call your attention to a recent work using a Mulliken-type definition of the bond order, which confirms also the bond order conservation ; Lendvay, J. Phys. Chem.)

SHUSTOROVICH -

1. The equilibrium distances r_1 and r_2 in $A^r_1 B^r_2 C$ have been calculated and, by comparing with the known values of r_{01} and r_{02} (for the gas-phase diatomics A^r_{01} and $B^r_{02} C$), the values of $X_1 = \exp [-(r_1 - r_{01}/a_1)]$ and $X_2 = \exp [-(r_2 - r_{02}/a_2)]$ have been determined. In all cases, the BOC at unity, $X_1 + X_2 = 1$, holds within ± 0.01 .

2. Thank you.

POTENTIAL MAPS OF CH₄, H₂O AND CH₃OH IN SILICALITE.
INFLUENCE OF THE SILICALITE STRUCTURE.

F. VIGNÉ-MAEDER

Institut de Recherche sur la Catalyse*, 2 avenue Albert Einstein, 69626 Villeurbanne
Cedex, France

ABSTRACT

The interaction energy of methane, water and methanol with silicalite has been evaluated by using an atom-atom potential containing atomic parameters, charges and dipoles deduced from ab initio calculations of small systems or fragments. Experimental geometry has been adopted for the silicalite framework and the adsorbed molecules which have been assumed to be rigid. The dependence of the potential values on the position of the atoms of silicalite has been studied by drawing the potential distribution corresponding to three different experimental structures.

I. INTRODUCTION

Zeolites are well-crystallised silicoaluminates which consist of silica SiO₄ tetrahedra, linked to each other by sharing all four oxygens. The isomorphous substitution of Si⁴⁺ by Al³⁺ creates an excess negative charge which is neutralized by metallic cations or protons. These materials present particular properties related to their microporous structure like large adsorption capacity, molecular shape selectivity, catalytic activity and they are widely used in industry. From an other point of view they are also ideal model systems for fundamental studies of microporous compounds because experimental data on structure or sorption are generally available and much of the methodology developed for biological systems should be directly transferable.

We have studied the interaction of small molecules with silicalite that is pure silica and corresponds to the end-member of the ZSM-5 series with a ratio Si/Al approaching infinity. It provides an example of a microporous adsorbent without any adsorption sites of chemical nature whose adsorption properties are only related to the channel structure. The framework of all the ZSM-5 zeolites is characterized by a network of straight channels interconnected by sinusoidal channels, both channels consisting of 10-membered oxygen rings.

The potential model that we used to calculate the interaction energy between silicalite and the adsorbed molecules CH₄, H₂O and CH₃OH is briefly described in section II.

From the interaction energies, a theoretical value of the initial heat of adsorption can be estimated by average over the orientations and positions of the adsorbed molecule, with the hypothesis of a Boltzman distribution. The experimental values are

* Laboratoire propre du CNRS, conventionné à l'Université Claude Bernard

satisfactorily reproduced but some regions of deep potential have been founded (ref.1). This is surprising for silicalite that does not contain any adsorption sites and for which rather uniform potential distributions could be expected. As this regions are close to the channel walls, we have examined the effect of the position of the atoms of the silicalite framework. We will present in section III some potential maps obtained by using the silicalite structures given by three different experimental works.

II. COMPUTATIONAL PROCEDURE

As usually the interaction energy has been expressed as a sum of four contributions, the electrostatic, polarization, dispersion and repulsion energy respectively. The interacting systems are assumed to be rigid so that each part can be represented as a sum of pairwise additive terms

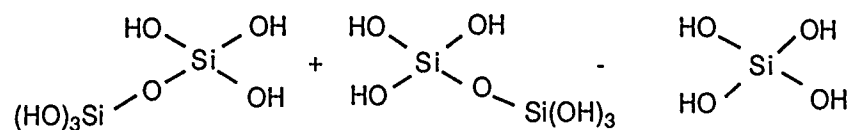
$$E = \sum_i \sum_j E_{ij}$$

where i and j denote the atoms (or other particular points) of the silicalite and of the adsorbed molecule respectively. As concerns the numerical evaluation, we have used a procedure developed by Claverie and coworkers (ref.2). The dispersion and repulsion terms are of Buckingham type (6-exp) and, to obtain the electrostatic and polarization parts, the abinitio charge distribution of each isolated system is represented as a multicentered multipole expansion.

The required atomic parameters for carbon, oxygen, hydrogen have been taken from previous works. The parameters for silicon have been calibrated with respect to the SCF interaction energy of the system disiloxane ($\text{H}_3\text{SiOSiH}_3$)-water at the equilibrium distance. All the SCF calculations have been performed with pseudopotentials and minimal basis sets.

In that procedure, the very large silicalite crystal should be replaced by a molecular system of reasonable size. Clusters with diameters of 24 Å appeared to be an acceptable compromise between accuracy and computer time. They contain about 200 silica tetrahedra whose free valences at the external surface are saturated with hydrogen atoms.

Nevertheless the size of these clusters is too large to allow SCF calculations. Their multipole representations have been reconstructed from those of dimers $(\text{HO})_3\text{SiOSi}(\text{OH})_3$ and monomers $\text{Si}(\text{OH})_4$ by superposition and subtraction of parts in excess (ref.3). For example, a trimer $(\text{HO})_3\text{SiOSi}(\text{OH})_2\text{OSi}(\text{OH})_3$ would be obtained by the scheme :



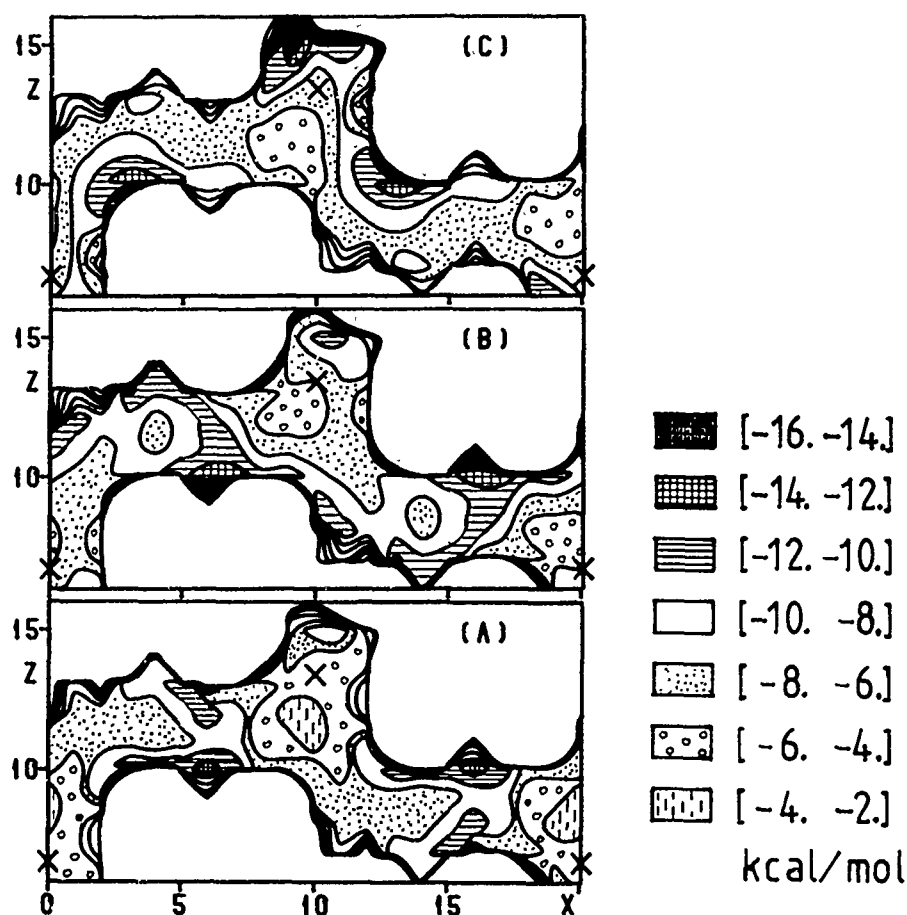


Fig. 1. Water-silicalite : potential map in the mirror plane of a sinusoidal channel ($y \approx 5 \text{ \AA}$) obtained with different crystallographic data for silicalite : (A) ref. 7, (B) ref. 6, (C) ref. 5.

Silicalite contains 12 different crystallographic sites for the silicon atoms, so that we had to perform SCF calculations for 12 monomers and 26 dimers of different geometries, which explains the use of pseudopotentials. The multipole expansions have been limited to charges and dipoles distributed on all the atoms. More details about the computational procedure are given in ref.1.

III. POTENTIAL MAPS

The framework structure has not been taken from the work of Flanigen on silicalite itself (ref.4) because some Si-O distances seem doubtful. But we used data on isostructural ZSM-5 zeolites. We have considered the results given by three different authors having used single crystal X-rays analysis and samples of different Al

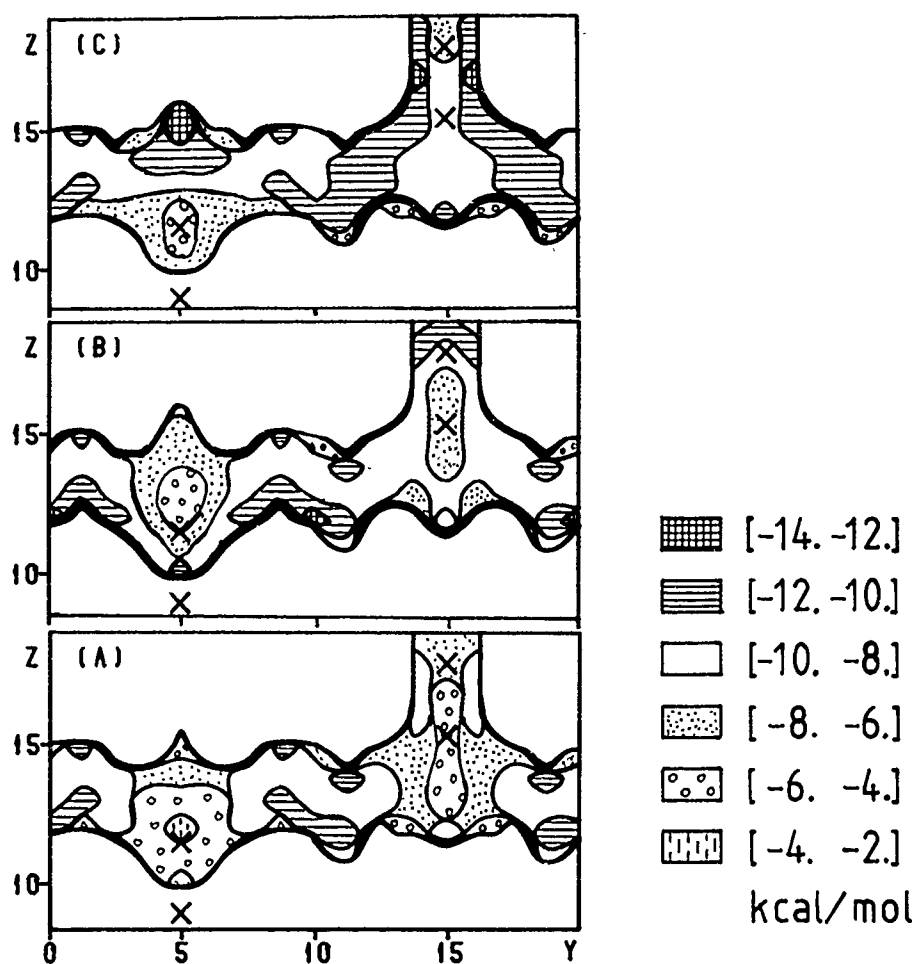


Fig. 2. Water silicalite : potential map in a longitudinal section of a straight channel ($x \approx 9 \text{ \AA}$) with different geometries for silicalite : (A) ref. 7, (B) ref. 6, (C) ref. 5.

content : 1,1 Al / unit cell for Olson (ref.5), 8.0 for Lerner (ref.6) and 0.3 for van Koningsveld (ref.7). All samples contained tetrapropylammonium ions, that are template ions required for growth and stabilization of the crystals, located at the channel intersection.

The general features of the straight and sinusoidal channels are given on figures 1 and 2 by the equipotential curves corresponding to the highest potential value (-2 kcal/mol for water). It is worthy to note that this curves do not delimit the channel void space but represent the positions of the water oxygen atom when the molecule is as close as possible to the channel wall. It is easy to observe along this curves the circular traces of the oxygen atoms constituting the channel wall. Both channels have roughly the same diameter and are interrupted by larger volumes at the intersections (for $x = 0, 10$ and 20 on figure 1 ; for $y = 5$ and 15 on figure 2).

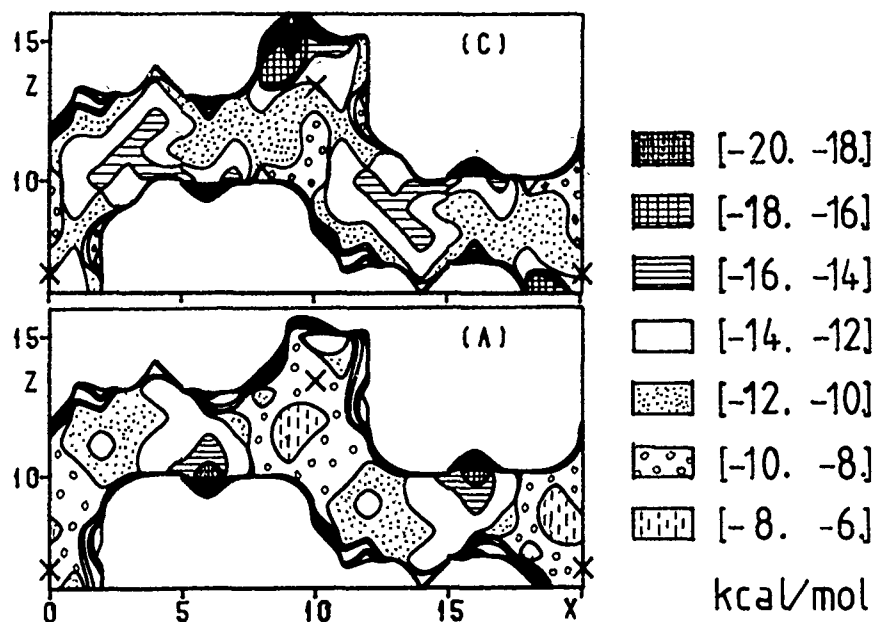


Fig. 3. Methanol-silicalite : potential map in the mirror plane of a sinusoidal channel. Silicalite structures taken from ref. 7 (A) and 5 (C).

The potential maps are plotted with the interaction energy values minimized with respect to the orientation. All the maps (Figs.1-4) present a domain of high potentials at the middle of the intersection (roughly - 3 kcal/mol for water and methane, -7 for methanol) where the molecule is most distant from the silicalite atoms. On the contrary in the channels the potential is deeper since the distances of the molecule with several framework atoms are smaller. The potential distribution is nearly uniform for methane because the main contribution is the dispersion term. It has been shown (ref.8) that in large cavities the interaction energy of dispersion type is deeper near the walls and tends to become uniform when the diameter of the cavity decreases up to the diameter of the adsorbed molecule. For water and methanol the electrostatic contribution that is sensible to specific charge environments yields a potential distribution in the channels more irregular than for methane. The maps for methanol looks like the addition of the maps of methane (the values of the dispersion part are very similar) and of water (similar values of the electrostatic part).

In table 1 are reported approximate values of the initial heat of adsorption obtained by average of the potentials over the orientation and the position of the adsorbed molecule, a Boltzman distribution being assumed. Our results are qualitatively in good accordance with the experimental ones. In particular the increasing order of CH_4 to CH_3OH is well reproduced. Our better values, corresponding to the structure of van Koningsveld (A), differ from the experimental ones by about 3 kcal/mol.

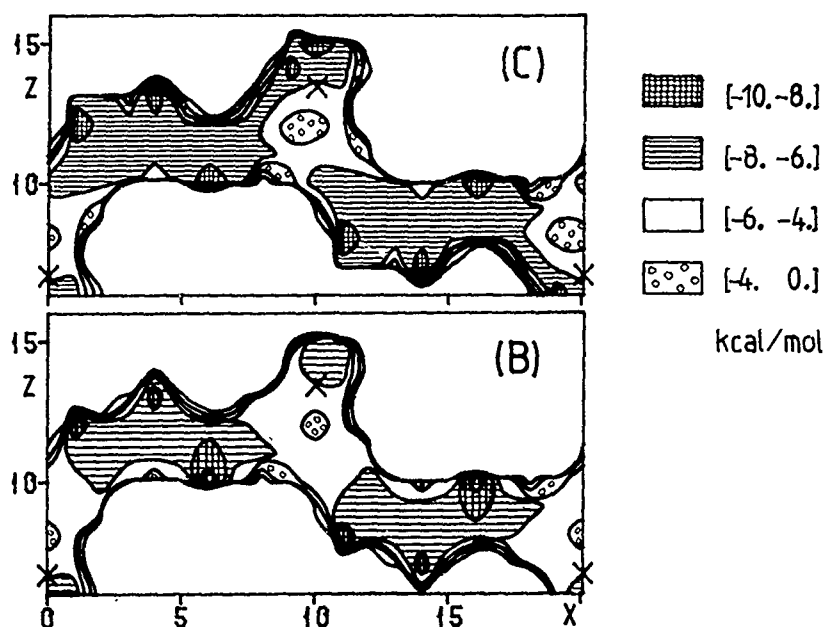


Fig. 4. Methane-silicalite : potential map in the mirror plane of a sinusoidal channel. Silicalite structures taken from ref. 6 (B) and 5 (C).

	Q_{theo} (kcal/mol)			Q_{exp} (kcal/mol)
	A	B	C	
CH ₄	9.4	11.	9.7	6.
H ₂ O	12.5	15.4	14.8	10.
CH ₃ OH	16.7		19.9	16.

TABLE 1

Adsorption heat Q_{theo} estimated from the potential values obtained with the silicalite structures of ref. 7 (A), ref. 6 (B) and ref. 5 (C). The experimental data Q_{exp} are average over several literature values.

The distribution of the potentials appear to be less satisfactory. Namely the potential maps show small regions of very deep potential that should correspond to strong adsorption sites. This is not corroborated by the experimental works in which the adsorption sites of the H-ZSM-5 zeolites are identified with the hydrogen atoms and should not be obtained in our model of silicalite. Potential holes are obtained with all the three silicalite structures, but they have different depths and positions. For example for water (Figs.1, 2) the deepest potentials are located at the channel intersection (-19.5 kcal/mol) for Olson's structure (C) and in the sinusoidal channel for both other structures (B) and (A) (-13.5 kcal/mol and -12.4 respectively). For methanol

(Fig.3) their positions are similar, but for methane the potential holes are more numerous and less deep (Fig.4). By examining the different contributions to the interaction energy it appears that all this large negative values of the potential correspond to a large electrostatic term.

We have more closely compared the potentials of water obtained with the structures (B) and (C) between which differences up to 4 kcal/mol have been obtained for the same positions and orientations of the water molecule. Combining with the structure (C) the charges and dipoles calculated with the structure (B) was not sufficient to reproduce the potentials fully obtained with (B). But they were satisfactorily reproduced by taking the structure (B) and the charges and dipoles corresponding to (C), so that the geometry adopted in the interaction energy calculations appears to be the determining factor. More precisely it is the position of only some framework atoms, close to the adsorbed molecule, that is very important. The potentials are sensitive to very small shifts in the atomic positions (not larger than 0.01 Å) because of the rather large charges carried by the silicon and oxygen atoms (about 2.7 and 1.3 a.u. respectively).

Thus the influence of the framework geometry on the potential distribution could be the result of the overestimation of the electrostatic part of the interaction energy. Further work will be needed to improve the quality of the multipole representation of silicalite, i.e. the quality of the SCF calculations on the fragments, for example by using larger basis sets without pseudopotentials and by taking into account the electric field created by the part of the crystal surrounding the considered fragment [ref. 9].

REFERENCES

- 1 F. Vigné-Maeder and A. Auroux to be published in J. Phys. Chem.
- 2 J. Caillat, P. Claverie, B. Pullman, *Acta Cryst.* B32 (1976) 2740 ; J. Langlet, P. Claverie, F. Caron, J.C. Boeue, *Int. J. Quantum Chem.* 20 (1981) 299 ; N. Gresh, P. Claverie, A. Pullman, *Int. J. Quantum Chem. Symp* 13 (1984) 243 ; F. Vigné-Maeder, P. Claverie, *J. Chem. Phys.* 88 (1988) 4934.
- 3 F. Vigné-Maeder, *Chem. Phys. Lett.* 133 (1986) 337.
- 4 E.M. Flanigen, J.M. Bennett, R.W. Grose, J.P. Cohen, R.L. Patton, R.M. Kirchner, J.V. Smith, *Nature* 271 (1978) 512.
- 5 D.H. Olson, G.T. Kokotailo, S.L. Lawton, *J. Phys. Chem.* 85 (1981) 2238.
- 6 H. Lerner, M. Draeger, J. Steffen, K.K. Unger, *Zeolites* 5 (1985) 131.
- 7 H. van Koningsveld, H. van Bekkum, J.C. Jansen, *Acta Cryst.* B43 (1987) 127.
- 8 E.G. Derouane, J.M. Andre, A.A. Lucas, *Chem. Phys. Lett.* 137 (1987) 336.
- 9 K. Seiti. These Université P. et M. Curie (Paris VI) 1988.

DISCUSSION

ANGYAN - 1. What are the charges of Si and O atoms in your model ?

2. The calculation of electrostatic interaction energy based on clusters of 24 Å edges may be misleading. In effect, these clusters are a little bit larger than one unit cell, which is in general not sufficient to have converged multiple lattice sums. These lattice sums are only conditionally convergent in the direct space. Standard techniques are available to make these lattice sums converge (Ewald, Bertaut) which should be used, in my opinion in the present case.

VIGNE-MAEDER -

1. Mean values : 2.7 for Si, 1.3 for O, 0.5 for H. These values depend slightly on the crystallographic site (variations lower than 5%).

2. This techniques should be actually used, at least in order to estimate the accuracy of our results. But they are very time consuming and not easy to implement. The small regions of deep potential that seem doubtful in our potential maps originate probably in local interactions and not in the truncation of the crystal.

MONQUE - Did you ever work or have you intended to model silicalite type zeolite or ZSM-5 with aluminium atoms included in the network ? Because as you probably know they are very used in catalysis.

VIGNE-MAEDER - We will probably consider such zeolites in the future. They are more difficult to model because they contain four types of atoms (Si, O, Al and cation) and the position of the aluminium atoms are not known experimentally. Their acidic properties have been up to now generally studied on small fragments of 2 to 5 Si or Al atoms and it will be indeed very important to introduce the influence of the framework. For catalysis, the modelling of silicalite is not uninteresting because it is well suitable for studying diffusion which represents an important part of the catalytic process.

Conformational studies on macrocyclic receptors and on their substrate complexes.

Georges Wipff

Institut de Chimie, 1, rue Blaise Pascal, 67000 Strasbourg (France)

SUMMARY

Several computer modelling techniques are used to provide deeper insights into the recognition properties of synthetic macrocyclic receptors, and to address the questions of preorganisation, complementarity, and binding selectivity. Alternatively, references to experimental data allows to outline present theoretical and computational limitations.

INTRODUCTION

Macrocyclic receptors synthesized and studied extensively since the pioneering works of C. Pedersen [1], J.-M. Lehn [2] and D.J. Cram [3] represent a fascinating field of research at the frontiers of chemistry, physics and biology. Among them, crown ethers, cryptands, spherands, podands and derivatives are able to act as receptors presenting a concave molecular region which bind strongly with marked specificities to small species such as cations, anions, neutral molecules [4-6].

We have been interested in modelling such receptors and their inclusion complexes because of their remarkable recognition properties, similar to those observed in large and complex biological systems. Recognition may extend beyond simple non covalent associations to processes which follow complexation, such as transport, or catalysis. It appears that in contrast to large biological species, much information is available concerning their constitution and structure(s) in the solid state, their spectroscopic properties in various solutions and environments, and thermodynamic and kinetic properties. Their relatively small size allows us to perform computer experiments in reasonable time. Finally, it appears that basic and practical questions raised in these modelling studies are in essence similar to those encountered with biological species and processes. We therefore felt it of importance to test and develop current modelling techniques, combining applied theoretical tools and high performance computer graphics. In the following, we illustrate uses of molecular mechanics (MM), molecular dynamics (MD), Monte Carlo (MC) simulations. Computer graphics will not be discussed as such, although being essential at both sides of the simulations, in providing input for programs, as well as for the visual analysis of the results.

CONFORMATIONS AND RECOGNITION

Given the recognition capabilities of these receptors, their structures and conformations are analyzed focusing on their ability to form "lock and key" complexes with potential substrates. Translation of this phenomenological scheme [7-10] into structural features implies that the conformation of the macrocycle is such that a cavity, or at least a concave region is present with a suitable orientation of the putative binding sites. Whether such a cavity is an intrinsic feature of the free receptor (in the "gas phase"), or if it is induced by the substrate, by a given solvent or environment are questions that are addressed using molecular modelling techniques. For instance, X-ray structures of 18-crown-6 (18-6) or of the bicyclic 222 cryptand in their free state and in the absence of molecular environment are such that they have no preformed cavity suitable for complexation [11, 12]. In such examples one would like to estimate the energy needed to organize cavities.

Classical conformational analysis describes primary processes such as inversion at atomic centers or rotation around bonds in molecules taking into account steric and electrostatic intramolecular interactions. For host/guest or receptor/substrate (R/S) complexes a "supermolecule" type approach involving the internal energy of R and S , their mutual interaction and the effect of the environment (e.g. counter ions, solvent molecules, etc..) on both R and S should be used.

The basic computational procedures require an adequate sampling of the conformational space of R , of R/S isolated or in their environment, and to calculate rapidly the energy of each set of coordinates. We will see examples and limitations of MM calculations on typical macrocycles in vacuo: monocyclic 18-6 [13], bicyclic 222 [14] and tricyclic SC24 [15] in vacuum (Fig. 1). To characterize in short the conformers, we refer either to their symmetry (e.g. 18-6 D_{3d}), or to the structure of the complex from which it has been extracted. For instance, SC24_N and 222_K are conformers of SC24 and 222 cryptands, extracted from the NH_4^+ and K^+ cryptates, respectively.

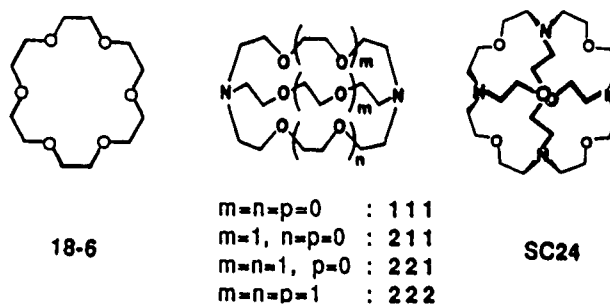


Fig. 1: 18-crown-6, the bicyclic and tricyclic cryptands.

Dynamic views of these flexible species and of the "lock and key" complementarity emerge from the analysis of their low frequency normal modes of vibration and from MD simulations [16]. Statistical analysis of structures found by high temperature

annealed MD simulations provides alternative pictures of the flexibility and conformational preferences [17]. Finally, modelling reaction paths for ion inclusion into compact cages illustrates the importance of flexibility/rigidity of their cage, and of the size of the ion on intrinsic barriers [18].

Significant theoretical progress has been made recently in the understanding and prediction of *binding selectivities in solution*. The method is based on the calculation of the change in free energy when one system is mutated into another via small perturbations [19]. The sampling of the conformational space at each step may be obtained by MC or MD simulations. We used an extended approach called the thermodynamic cycle perturbation method and determined relative binding energies. This procedure was tested to account quantitatively for the Cl⁻/Br⁻ binding selectivity of **SC24,4H⁺** [20]. It has been used recently in the macrocyclic area to calculate the relative binding free energies of Na⁺/K⁺ by **18-6** in water [21] and in methanol [22], of nitromethane/ malononitrile/ acetonitrile by **18-6** [23], or of Na⁺/K⁺ by dibenzo-crown ethers [24] as well as the binding of pyridine/pyrazine by one of Rebek's acridine diacid receptors in chloroform [25]. More generally, this promising technique should lead to relative free energies of solvation for closely related species or conformers, or of binding involving biological receptors [19]. The studies reported so far deal however with systems whose structure and binding properties are known experimentally. Although being one of the ultimate goals of molecular modelling, quantitative prediction of binding selectivities is still a formidable task, mainly because of conformational sampling and energy representation problems.

ENERGY REPRESENTATION AND COMPUTATIONAL TECHNIQUES

Molecular mechanics, and dynamics calculations were performed with the program AMBER [26], using the representation of the potential energy given in reference 27.

The bonds and bond angles are treated as harmonic springs, and a torsional term is associated to the dihedral angles. The interactions between atoms separated by at least three bonds are described within a pairwise additive scheme by a 1-6-12 potential.

The parameters are derived from the AMBER force field [27] and can be found in references 13-16. Most critical is the electrostatic representation of the system, which was as much as possible calibrated from gas phase data. The lack of an explicit polarization term in this force field led us to use different sets of charges depending of the presence and nature of host ion. For the uncomplexed macrocycles, the charges ($q_O = -0.3$, $q_N = -0.24$) account for the dipole moment of the OMe₂, NMe₃ building units. For the alkali cation complexes, larger values ($q_O = -0.6$, $q_N = -0.6$) were required to reproduce experimental M⁺...OMe₂ interaction enthalpies in the gas phase. A similar procedure was used to derive charges involving ammonium sites or the NH₄⁺ substrate. Explicit account of distance and orientation dependent polarization energy and of many body effects would be more satisfactory [28], but this is not incorporated in

current modelling software performing MD simulations with reliable parameters, and requires extensive computer time.

MM optimisations relax the starting structure to the nearest energy minimum and have to be performed consistently in terms of force field and procedure. We used conjugate gradient minimization followed in most cases by a Newton Raphson optimization to check for the second derivatives of the energy, and to obtain the normal modes of vibration.

MD simulations were run for 100 ps at 300K starting with random velocities and using the Verlet algorithm with a time step of 1 fs [26].

BUILDING SELECTED 3D STRUCTURES

Finding computationally the 3D cartesian coordinates of macrocycles is a very challenging problem because of their (poly)cyclic nature and of the way the structural requirements are formulated. For instance, one would make the complex of **18-6** with ammonium derivatives (Fig. 2a) such that the crown anchors the NH_3^+ moiety, that the lateral substituents are in axial position, and that the ester carbonyl of the peptidic substrate may be attacked by one sulfur of the receptor.

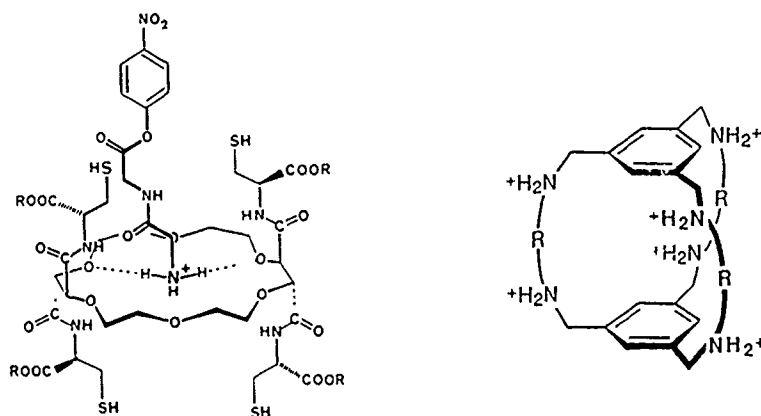


Fig. 2a: (left:) "supermolecule" formed between a lateral derivative of **18-6** and an ammonium substrate (from reference 5). Fig. 2b : (right): the "Barrel,6H⁺" receptor ($\text{R} = \text{CH}_2\text{-CH}_2\text{-CH}_2$) of NO_3^- (from reference 67).

Classical methods based on incremental building in terms of internal coordinates cannot be used because dihedral angles are unknown. Manipulations of 3D structures done easily in the hands of the chemist with CPK models [29], involving concerted rotations around several bonds in such a way that a cavity forms, keeping the connectivity, and such that the binding sites are properly oriented cannot be done with current modelling software, especially for polycyclic systems. On the other hand, building methods such as symmetry repetition of given fragments [30] or drawing

structures within a diamond lattice [31] are specific and cannot be used for general purposes.

The Crippen's approach [32] based on the partially known matrix of interatomic distances revealed to be successful in translating constraints such as range of distances between non bonded centers or of dihedral angles into distances. From the upper and lower bounds of that matrix, several sets of coordinates could be calculated [33]. The 2D3D graphics interface allowed us to input graphically the data from 2D drawings, and to visualize and analyze the 3D structures generated [34,35]. For instance, "in-out" forms of **222** (Fig. 3) were generated by imposing Cc-N-LP angles of

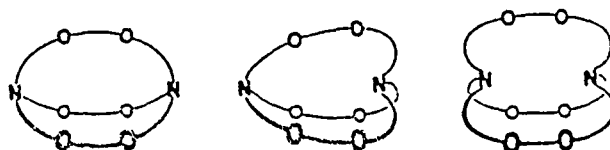


Fig. 3 : "out-out", "out-in" and "in-in" topomers of the bicyclic **222** cryptand.

0° and 180°, respectively (Cc is the Center of the cavity, and LP is a point along the nitrogen lone pair) [14]. Similarly, coordinates of various "in"/"out" topomers of **SC24** could be obtained [15]. One major limitation of this approach is the small number of different structures produced and the bias introduced by the constraints. In addition, in order to find solutions a balance between short range and long range constraints is required.

The ELLIPSE technique of Billeter et al. [36, 37] like DISMAN [38] uses the torsional angles as variables and was able to build selected forms of highly connected and constrained species, such as the "Barrel", $6\text{H}^+/\text{NO}_3^-$ inclusive complex (Fig. 2b) [39]. Here, in addition to the ring closure requirements, the constraint was to form H-bonds involving the twelve N-H^+ and NO_3^- . However, such purely geometrical techniques as ELLIPSE or distance geometry may not lead to low energy forms, nor to the absolute minimum. They have therefore to be combined with energy optimizations and structure relaxations [40]. P.A. Kollman reported such a combined use of ELLIPSE and MD simulations for **18-6** [37]. Houk et al. modelled phenanthrene macrocycles using ELLIPSE and AMBER or MM2 optimisations [41]. We performed a conformational analysis of **18-6**, **222**, **SC24** neutral and protonated combining distance geometry searches with AMBER [13-15].

Monte Carlo searches taking the dihedral angles as variables [42] can hardly be used for polycyclic systems, especially when additional structural criteria have to be met. The Saunders stochastic method [43] combines random atomic displacements with energy refinements, sampling largely the conformational space of (poly)cyclic hydrocarbons and "hunting" for the global minimum. The latter method might be used for macrocyclic receptors as well. Annealing procedures used in conjunction with

Monte Carlo simulations [44] have been reported for open or monocyclic molecules, but have not been tested yet on polycyclic systems.

Finally, quenched or annealed MD simulations at high temperature incorporating constraints should also provide suitable structures, as used for modelling structures from NMR data [45].

MOLECULAR MECHANICS OPTIMIZATIONS "IN VACUO" ENERGY COMPARISON OF X-RAY STRUCTURES.

Structures in the crystalline state [11,12] may differ from those in solution, but are precious as a proof of the inclusive nature of the complexes, and of their conformational flexibility. For instance, **18-6** is either of C_i , D_{3d} or C_1 symmetry when respectively uncomplexed and in the absence of molecular environment, complexing K^+ or NH_4^+ , or complexing Na^+ . The **222** cryptand has an "in-in" orientation of the nitrogen bridgeheads in all structures, except in the bis- BH_3 adduct which is "out-out", and will be noted hereafter **222_OO**. The free **222** (noted **222_II**) is quite elongated, whereas its Na^+ , K^+ , NH_4^+ , Cs^+ , Tl^+ cryptates have a similar D_3 form with converging N,O binding sites. With Ag^+ or Ca^{2+} as hosts, **222** is distorted with a smaller cavity. No "in-out" form has been found in crystals, although probably present in protic solvents [2]. For **SC24**, the two X-ray structures available $NH_4^+/SC24$ and $Cl^-/SC24, 4H^+$ are "in-in-in-in" with similar cavity sizes, but correspond to different conformations [11]. Since the structure of **SC24** free could not be solved it is not clear from experiment whether **SC24** has a preformed cavity, or if the cavity is induced upon complexation. However modelling studies indicate that various "in"/"out" topomers of **SC24** uncomplexed have a quasi tetrahedral cavity of fairly constant size [15].

Energy comparison of these forms, although being in principle a straightforward procedure suffers limitations related to the force field and to the optimisation procedures. Comparison of MM and of ab initio calculations on **18-6** in its C_i , D_{3d} and C_1 conformations illustrate the first point [13, 37, 46]. There is agreement on the fact that the distorted C_1 form is the least stable, but not on the relative C_i/D_{3d} stabilities. The discrepancy is mostly related to the electrostatic effects: the repulsive 1-4 OC-CO interactions in the gauche arrangement disfavor D_{3d} relative to C_i . Similarly, relative conformational energies of **222** obtained by AMBER [14] and MM2 [47] differ. It is therefore important to run several sets of calculations varying critical parameters such as the dielectric constant or the atomic charges in order to assess the origin and range of conformational preferences.

Relative stabilities and optimized structures depend also on the minimization procedure. For instance, using the conjugate gradients algorithm gave increased stabilities for the **Ca**, **Na**, and **K** conformers of the **222** cryptand, in qualitative agreement with the expected strain induced in the Ca^{2+} , Na^+ and K^+ cryptates, respectively. However, Newton-Raphson optimizations converged to a same minimum of D_3 symmetry [14].

In some optimizations, we wanted to keep the "optimized" structure as close as possible to the experimental one in order to prevent similar structures converging to the same minimum, and to compare the deformation energies induced by complexation. We found with the **222** cryptand that no conclusive results could be obtained by constraining the coordinates to their initial values because the results depend on the choice of the restraining force constant. On the other hand, experimental structures have to be relaxed because of inhomogeneities in bond lengths due to crystal disorder, or to thermal effects in the crystal (for instance, the C-O bonds of **18-6** are on the average 1.36 Å at 300K, and 1.42 Å at 100K [48]).

Finally, concerning this search for energy minima it is worth mentioning that whereas pieces of macrocycles extracted from experimental structures are close to local energy minima, modelbuilt ones are not and need more extensive relaxation, combining for instance MM and MD simulations.

MM optimisations of **222** and **18-6** uncomplexed confirmed that their free forms found in the solid state (respectively **222_II** and **C_i**) are more stable than those extracted from the complexes, which indicates that these structures do not result from significant packing forces. However, the energy difference is weak (of a few Kcal/mole) compared to the interaction energy with an ion (respectively about 95 and 115 Kcal/mole, for the **18-6** and **222/K⁺** complexes). In other words, complexation induces a structural reorganization, but with no significant strain, especially for the best "recognized" cation in solution (**K⁺** for **18-6** and **222**). Unlike crown ethers and cryptands, anisole spherands [49] display little conformational flexibility, and become preorganized for complexation during their chemical synthesis [3, 6].

Optimisations of the alkali cation complexes of **18-6** in various conformations were able to account for their structures: **Na⁺** prefers the **C₁** form, **K⁺** the **D_{3d}** form in a nested position, and **Cs⁺** a perched position over the **D_{3d}** ring, as found in the crystal [13]. Similary optimisations of the **222/K⁺** cryptates starting from **222_OO** or **222_II** forms converged to the **K** form because of the electrostatic and steric strain induced by **K⁺**. With **222** however, no **K⁺/Na⁺** discrimination could be found among the **K/Na/Ag** conformers [14].

In these cation complexes, we calculate a decreasing complexation energy from **Na⁺** to **Cs⁺**, as observed in the gas phase for binding of these ions to ether or amine binding sites. In aqueous solution, however, the selectivity peaks experimentally at **K⁺** for **18-6** and **222**, at **Rb⁺** for **SC24**, and at **Na⁺** for the anisole spherand [2, 5, 6]. Similar peaks are calculated when one subtracts the experimental hydration enthalpies of these ions from the calculated complexation energies. However, such a naïve procedure is unable to account for the binding selectivity of one given ion to different receptors [16,34] and could not be used reasonably for prediction purposes.

Fragments of experimental structures were used to model ammonium complexes of **18-6** and of its lateral amide derivatives [49]. MM optimisations account for the binding selectivity (primary > secondary > tertiary ammonium) observed in the gas phase and in

solution, as well as for the different modes of binding. However, when the structural complexity of the system increases, e.g. with longer ammonium substrates or lateral arms, no conclusion could be drawn because of the multiple minima problem [49]. Particularly, the ability of that system to display chiral recognition could not be determined.

Chiral recognition of ammonium substrates was studied with the chiral Cram's receptor based on binaphthyl units (Fig. 4) and its L/D complexes with (Phenyl)-⁺Gly(OMe) [49]. The complex with the L substrate was built from the X-ray structure of the macrocycle with the D substrate. Here, the MM optimisations of the complexes account for the weak energy difference (2 Kcal/mole) involved in the chiral recognition displayed by this macrocycle.

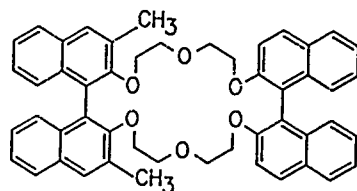


Fig. 4 : The Cram's chiral macrocycle.

HARMONIC DYNAMICS. LOW FREQUENCY NORMAL MODES OF VIBRATION

The low frequency normal modes of vibration give insight into molecular deformations of low energy (less than 100 cm^{-1}) and of quite large amplitude. Atomic displacements of about 3 \AA can be achieved at less than 5 Kcal/mole. The display of these vibrations on the PS300 graphics system using MDNM, [50] either with static vectors of elongation, or as dynamic pictures, revealed interesting qualitative features which may be related to the recognition properties of these macrocyclic receptors [16].

The type of motions: qualitative features.

First, for the free receptors, from the monocyclic 18-6 to the tricyclic SC24, among the three first vibrational modes, a particular one is found which leads to an opening of the cavity, and makes its binding sites more accessible to the solvent, or to an approaching substrate (Fig. 5) In other words, *there are normal modes of vibration of the free receptor which provide a path for substrate inclusion*. This feature is not restricted to the conformers extracted from cation complexes. For instance for 18-6, V_7

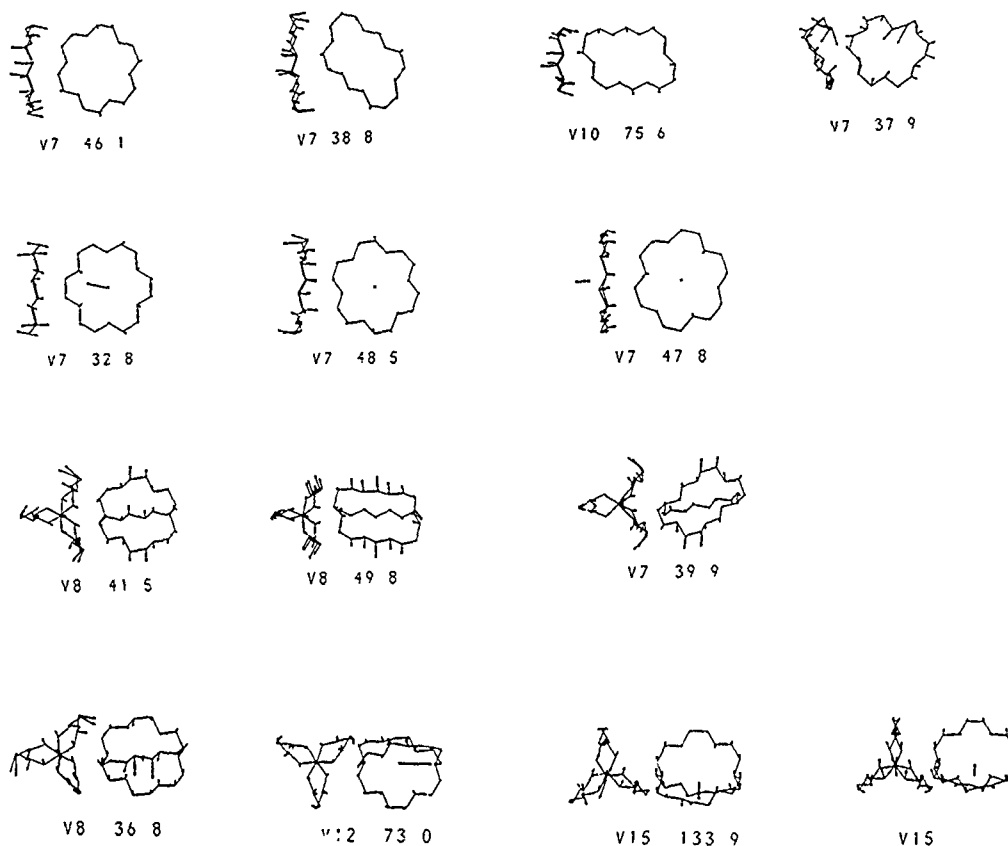


Fig. 5 : Typical low frequency normal modes of vibration (orthogonal views) of 18-6 free (line 1) in its D_{3d} , C_i and C_1 forms, of 18-6 Na^+ , K^+ , Cs^+ complexes (line 2), of 222_K, 222_II, 222_OO free (line 3), of 222/ M^+ (respectively K^+ , Na^+ , K^+ , Cs^+) cryptates (line 4).

and V_8 of the D_{3d} form, as well as V_7 and V_{10} of the C_i form, or V_7 and V_9 of the C_1 form correspond to a same type of ring folding. For the bicyclic 222 cryptand, V_8 and V_9 of the 222_K form, V_7 and V_8 of the 222_II form, or V_7 of the 222_OO form are such that two bridges open/close and make the cavity less shielded. In SC24 neutral, or tetraprotonated SC24,4H⁺ in the "in-in-in-in" form, one face of the tetrahedron formed by the nitrogens opens while the opposed nitrogen approaches, making a kind of breathing motion of C_3 symmetry.

In a barrel type tricyclic topology, the "Barrel,6H⁺" (Fig. 2b) receptor of NO_3^- [67] displays opening motions of the ammonium chains among the first vibrations. Such molecular motions are probably a general feature of receptors, which have been characterized in larger biological systems like lysozyme, trypsin [51].

In the complexed macrocycles, similar vibrations are found, which provide paths for cation extrusion and solvation. See for instance v7 for the 18-6/ K^+ complex. For the

222/ Na^+ , K^+ and Cs^+ cryptates, V_8 is the characteristic mode (Fig. 5). Similarly, V_8 of the **SC24**/ K^+ or **SC24**/ NH_4^+ cryptates is a "decomplexation mode" [34].

Vibrations of the ion inside the cages reflects also the ion/cage complementarity. Indeed, when the cation fits into the cage, it is immobile in the lowest frequency modes; otherwise it moves in a characteristic way. For instance in V_7 of the **18-6** $_{\text{D}3\text{d}}$ complex, Na^+ , which is slightly too small undergoes in plane librations of large amplitude; Cs^+ which is too large oscillates along the C_3 axis above the ring, whereas K^+ which has the right size does not move (Fig. 5). In the **222** cryptates, the first motion for Na^+ is a large oscillation along the $\text{N}\cdots\text{N}$ axis inside the cage (in V_{12} , 73 cm^{-1}), whereas K^+ is expelled through one face at higher frequency (V_{15} , 134 cm^{-1}) and Cs^+ has no significant motion.

Vibrational frequencies.

Quantitatively, we found these low frequencies to be rather insensitive to the electrostatic representation of the molecule. The vibrational spectrum becomes more rich in low frequencies from **18-6**, to **222** and **SC24**. The reverse order might have been anticipated based on the intuitive idea that increased connectivity and rigidification would increase the vibrational frequencies, but the results simply follow the number of atoms and dihedral angles.

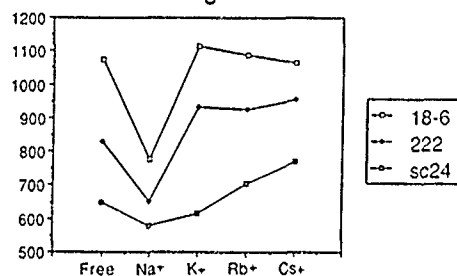


Fig. 6 : Sum of the 10 lowest frequencies of vibration (in cm^{-1}) for **18-6** $_{\text{D}3\text{d}}$, **222** $_{\text{K}}$, **SC24** $_{\text{N}}$ free and in their Na^+ , K^+ , Rb^+ , Cs^+ complexes.

Upon cation complexation, the receptor should be rigidified by the electrostatic and steric strains induced by the cation. Fig. 6 displays the sum (S_{10}) of the, for example, ten lowest frequencies for the free and complexed macrocycles. In all the complexes of **18-6**, **222** and **SC24**, S_{10} is higher for Cs^+ than for Na^+ . There are however two interesting observations related to the cation/crown complementarity. First, S_{10} is smaller for Na^+ complexes, than for the free macrocycles, probably because Na^+ is too small and undergoes large amplitude librations within these cages. Second, S_{10} depends on the size and flexibility of the cage, going from Na^+ to Cs^+ . **SC24** is rather rigid with fixed dimensions [15], and the lowest frequencies increase regularly. A similar trend was observed for the Li^+ , Na^+ and K^+ complexes of a rigid "preorganized" anisole spherand [49]. **222** is more flexible and adjusts itself around the cations: going from K^+ to Cs^+ a plateau is observed. **18-6** makes with Na^+ and K^+ inclusion type complexes, but Cs^+ , too big, sits over the ring making a weaker complex, and the low frequencies decrease from K^+ to Cs^+ . To determine to which extent these differences in

dynamic behaviors are specific to these macrocycles needs further investigations. Nevertheless we feel that they express in some way differences in the receptor/substrate complementarity and in flexibility.

MOLECULAR DYNAMICS SIMULATIONS AT 300K

Such simulations provide alternative insights into conformational stabilities of the free and complexed receptors, into the flexibility as a function of the conformation of the macrocycle and of the nature of the complexed ion, i.e. into the *R/S* complementarity [16]. The Root Mean Square atomic displacements (RMS) from the average position during the MD simulation were calculated to size the amplitude of the motions.

First, in contrast to what we observed for the normal modes of vibration, there is a clear effect of the electrostatic representations on these RMS's. Particularly the free macrocycles simulated with charges of low polarity ($q_O = -0.3$) retain their starting X-ray conformation, but undergo conformational changes with large fluctuations when polar charges are used ($q_O = -0.6$). This is because gauche OC-CO arrangements become destabilized relative to trans by 1-4 repulsive O...O interactions.

The uncomplexed macrocycles.

Comparison of the RMS's for uncomplexed **18-6**, **222**, **SC24** in several conformations and for **111**, **211**, **221** to **222** in the bicyclic series demonstrate the *effect of the conformation on the molecular mobility*. Again, considering the connectivity criteria alone, one might have expected to find decreased mobility and RMS's from **18-6** to **SC24**. In the bicyclic series, rigidification might similarly be expected going from **222** to **111**. The calculated RMS's averaged by atom groups, or over the whole molecules show that it is not so. Among the conformers with the lowest RMS's, one finds the order bicyclic **222_II** (0.32 Å) < monocyclic **18-6_D3d** (0.34 Å) < tricyclic **SC24_N** (0.47 Å) ! The largest RMS's are found for the three types of topologies, in the order **SC24_CI** (0.53 Å) < **222_K** (0.54 Å) < **18-6_C1** (0.56 Å). In the bicyclic series, the **111** and **222** cryptands fluctuate respectively by 0.23 and 0.32 Å. As expected, the "larger" molecule fluctuates more than the "smaller" one. Nevertheless, as an effect of topology, size and conformation **221_K** shows a singularly larger fluctuation (0.56 Å)

How can relative mobilities be rationalized? They might seem first to depend on the relative conformational stabilities. For **18-6** and **222**, the most stable conformers (respectively **C1**, close to **D3d** and **222_II**) have indeed the lowest fluctuations. Conversely, the **C1** form of **18-6** which has the largest RMS is highest in energy and **SC24_CI**, less stable than **SC24_N** fluctuates more (respectively of 0.53 and 0.47 Å). There are however exceptions like **222_K**, which is more mobile than **222_OO** (respectively of 0.52 and 0.44 Å), although being more stable by 5.4 kcal/mole. Similarly, among the three forms of **221** extracted respectively from its Co^{2+} , Na^+ and K^+ cryptates, **221_Co** is the least stable compared to **221_Na** or **221_K** (respectively 4.4, 0.5 and 0.0 kcal/mole). But, the order of mobilities is reversed (respectively

0.33, 0.47 and 0.56 Å). Here, the order follows the size of the cavity left by Co^{2+} , Na^+ and K^+ (respectively 2.22, 2.65 and 2.91 Å for the cation...N distances, and 2.16, 2.28 and 2.80 Å for the cation...O distances [11]). This relation seems to be an interesting clue. The *cavity size effect*, according to which *conformers with largest empty cavities display the largest fluctuations* may also rationalize why elongated **222_II** is significantly less mobile than **222_K** extracted from the K^+ cryptate. The cavity of **SC24** is larger than that of **222** and its fluctuations are larger, although **SC24** is more connected. Concerning its two forms, **SC24_CI** extracted from the **SC24,4H⁺/Cl⁻** cryptate is more mobile than **SC24_N** extracted from the **SC24/NH₄⁺** cryptate, although both have similar cavity sizes. This may be presumably related to the loss of the organizing effect of the four protons and of the anion in the former. In more general terms, it appears thus that *the conformer of the receptor induced by substrate binding has a large mobility in the absence of the substrate* and that *free macrocycles tend to fill their cavities* [41]. It is stressed that in that respect, the solvent will play a particular role.

The cation complexes.

In the cation complexes of **18-6**, **222** and **SC24** calculated with polar set of charges, one observes first a *rigidification* compared to the free receptor. For the K^+ cryptates of **18-6**, **222_K** and **SC24_N**, the RMS's drop respectively by 0.21, 0.40 and 0.12 Å. There is in addition a clear *effect of cation/receptor complementarity on the relative motions of the cation and of the cage*. Indeed, from Na^+ to Cs^+ , the fluctuations of the cage decreases in the three macrocycles: from 0.33 to 0.28 Å in **18-6**, from 0.40 to 0.23 Å in **222** and from 0.33 to 0.29 Å in **SC24**, as a result of the increased size of the ion. Although Na^+ interacts more strongly with the cages than K^+ or Cs^+ , it does not rigidify them as much.

It is interesting to compare the relative mobility of the cations within the **222** and the **SC24** cryptates. In general, it decreases as the size of the ion increases but Cs^+ which is somewhat more compressed in **222** than in **SC24**, fluctuates slightly less (respectively 0.14 and 0.16 Å). A dramatic difference is observed for the smallest ion. Na^+ fluctuates in **SC24** (0.55 Å) more than twice as in **222** (0.25 Å)! We believe that this is a result of the difference in flexibilities of these cages: **222** is flexible not only in its ability to adopt various conformations, but also in being able in a given conformation to adjust its size to that of the substrate. On the other hand, the more connected **SC24** is more rigid and keeps a similar cavity size in its free state and in its inclusion complexes, but that cavity is too large for Na^+ . In this context, it is also noteworthy that in the optimized **Na⁺/222** cryptate, Na^+ is at the center and equidistant from the binding sites whereas in the **Na⁺/SC24** cryptate Na^+ moves from the center to make closer contacts with three oxygens of the cage [15]. The flexibility of **222** reaches however a limit in the Ca^{2+} cryptate, in which the $\text{Ca}^{2+}\cdots\text{C}$ distances are longer than the sum of van der Waals radii [11].

The alkali cations complexes of **18-6_D3d**, in contrast with those of **222** and **SC24**, display an interesting non regular behaviour from Na^+ to Cs^+ . This is because Na^+ and K^+ can fit into the ring, whereas Cs^+ too big for **18-6** sits and fluctuates over the ring. As a result, Cs^+ is less anchored than K^+ and fluctuates more, respectively by 0.28 and 0.24 Å. For **18-6**, the RMS's peak at Rb^+ , whose size allows fluctuations between "inclusive" and "perched" type positions.

MD simulations on ammonium complexes of **18-6** free and substituted by lateral amide arms [52], modelled as anchoring site in an "artificial enzyme" shed light on the *possible importance of the mobility for efficient catalysis* by such systems. Indeed, we calculate that amidic fragments increase both the stability and the mobility of R-NH_3^+ complexes compared to **18-6** unsubstituted. It might have been anticipated that increased stability induces reduced mobility. Above, for the alkali cation complexes of **222** and **SC24**, we have seen that it was not so. In the ammonium complexes of lateral amide derivatives of **18-6**, the electric field of the amide carbonyls facilitates indeed both the binding of the substrate and its oscillations away from the crown ether oxygens. Catalytic behavior requires the supermolecule to be flexible enough to move from stable non covalent "Michaelis complexes" to stabilized transition states involving covalent binding between reactive centers of the receptor and of the substrate [52,55]. Our MD results confirm that such an effect operates in "artificial enzymes based on macrocyclic units in which *enhanced stability is accompanied by increased mobility* [52]. Biological catalytic systems might take advantage of that dynamic effect as well.

SAMPLING OF THE CONFORMATIONAL SPACE USING HIGH TEMPERATURE ANNEALED MD SIMULATIONS.

In the MD simulations at 300K, run for 100 ps, no significant conformational change took place for the **18-6**, **222** and **SC24** complexes. For the free **18-6_C1**, unsymmetrical forms appeared in 400 ps, none of them being more stable than **18-6_C1**. We paid a particular attention in the bicyclic series to the **222** cryptand and to the occurrence of nitrogen inversion at the bridgeheads. In fact, starting from the **222_OO** form, no inversion occurred within 500 ps using several N-C-N force constants, whereas such "out" to "in" conversion takes place rapidly in solution [2]. It involves concerted reorganization of the bridges, which does not take place either in these MD simulations. Whether this results from the too short simulation time, or from the lack of solvent molecules in these calculations has to be assessed by appropriate computer experiments.

We tested with the **222** cryptand an alternative procedure to interconvert structures more efficiently, and sample the conformational space, i.e. High Temperature Annealed MD simulations [17, 56]. First the molecule is severely shaken by MD at high temperature (1000K) for 100 ps, then each of the 500 structures saved every 0.2 ps is energy minimized by MM, reshaken at 300K during 20 ps of MD in order to allow for enough relaxation, and finally reoptimized.

Starting from several structures, including **222_OO**, it was gratifying to find in the final collection of 500 annealed structures the experimental **222_II** one which was calculated as being the most stable among X-ray conformers, as well as new "in-in" slightly more stable ones. Thus, this procedure is able to produce "the" global minimum and confirms that several forms similar in shape (quite elongated and without cavity) are in equilibrium.

The disappointing result is the *absence of any form of cryptate* in that set, and particularly of the **K** conformer which displays the highest complementarity for K^+ , the alkali cation which is the best recognized by **222**. Several simulations were repeated including a gradual representation of the complexed cation, either as a pure electrostatic driver (+1 dimensionless charge) or with an electrostatic + steric representation which simulates K^+ . We found that the **K** form was generated only in the presence of K^+ , although being only a few kcal/mole higher in energy than **222_II**. This shows that particular forms of **222** acting as receptors and ionophores, like those which bind Ca^{2+} , Ag^+ , or Pb^{2+} have a low probability to be present in the absence of their substrate, and that **222** is *topologically, but not conformationally preorganized for cation complexation*. More generally, in terms of drug design, this suggests that conformations of a flexible drug able to bind to a receptor have very low probability to be found computationally if the supermolecule formed with the receptor is not considered explicitly.

During the MD run at 1000K several "in"/"out" conversions take place, and the final set of annealed structures contains populations of "in-in", "in-out" and "out-out" topomers. The energy distribution of these classes shows a clear preference for converging orientations of the brigheads. The populations increase in the order "out-out" < "in-out" < "in-in" (respectively 7%, 33% and 61%) and the peaks of these classes (respectively at 10.1, 8.7 and 6.2 Kcal/mole from **222_II**), as well as the lowest energy forms (respectively at 4.4, 3.4, -0.8 kcal/mole from **222_II**) confirm the preference for "in" forms.

New minima of each topology are also found. Two "in-in" forms are slightly lower than **222_II**, three "out-out" forms are more stable than **222_OO** extracted from a molecular adduct [11], and there are several "in-out" forms at less than 4 kcal/mole from **222_II**. The most stable "in-in", "in-out" and "out-out" forms are shown in Fig. 7.

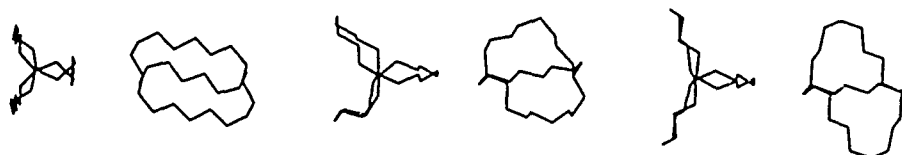


Fig. 7 : Lowest energy conformers of **222** "in-in", "in-out" and "out-out" generated by the High Temperature Annealed MD simulations.

Interestingly, most of low energy forms have *non converging* orientations of the oxygen binding sites. Several have indeed a large "solvent accessible surface" [57], comparable to that of a thia analogue of **222** whose heteroatoms point outside [58]. Such conformers, not characterized in the solid state for **222** itself, should be significantly better hydrated than the hydrophobic ones extracted from cation inclusion complexes, and play a particular role for ion capture, as well as for the ionophoric behavior. It is also worth noting that the energy difference between stable "in" and "out" forms is small enough to be compensated by hydrogen bonds involving a protic solvent, and therefore these forms are expected to be in equilibrium in solution.

REACTION PATHS FOR INCLUSION OF THE SUBSTRATE INTO THE RECEPTOR

Conformational analysis of macrocyclic supermolecules should not only consider the extreme states (free versus complexed macrocycle), but also the process of ion capture. The energy and structure of transition states for ion inclusion as a function of the substrate is of particular interest. Experimental data for complexation of alkali cations by cryptands in solution show that complexation is fast compared to decomplexation, and driven by the stability of the inclusion complex formed [59].

Modelling the complexation process in solution is a difficult task due to the conformational changes of the receptor, and to desolvation/solvation effects. Recently, Pohorille et al. modelled the capture of Cl^- by **SC24,4H⁺** in water using MD simulations and found that desolvation of Cl^- and deformation of the receptor contribute mostly to the barrier [60].

Reaction path and kinetics for complexation of Na^+ by **18-6** in vacuo has been calculated using canonical variational transition-state theory [61].

We modelled the inclusion of alkali and NH_4^+ cations into **222** and **SC24**, and the inclusion of halides into **SC24,4H⁺** in the gas phase [15, 34], using step by step MM optimisations and MD simulations. Such calculations although being crude shed light on intrinsic features influencing the barrier: size and charge of the ion, and flexibility of the cage. Indeed, no barrier is calculated for the inclusion of Na^+ to Rb^+ or of NH_4^+ into **222** and **SC24**. The Cs^+ case appears to be critical. Cs^+ enters into **SC24** with a 5 Kcal/mole barrier, but with no barrier into **222**, despite the fact that the cavity of **SC24** is somewhat larger. This "gas phase" kinetic effect results mostly from the rigidity of **SC24** compared to **222**: the passage of Cs^+ through a face of **SC24** induces more strain and deformation energy

For the Cl^- and Br^- anions captured by **SC24,4H⁺**, calculated energy barriers of 7.6 and 12.5 Kcal/mole separate external and inclusive complexes "in vacuo". Whereas for Cl^- , the inclusive form is preferred over the exclusive one by 6.9 kcal/mole, for Br^- , both forms have similar stabilities due to the strain induced by Br^- in the cryptate. The nature of the complexes in solution is not clear, because they have on the NMR time scale an average tetrahedral symmetry [62]. However, given the relative stability constants for Cl^- and Br^- complexes (the Log K_s values being respectively > 4

and < 1 in water) and the results of the above simulations, it may be reasonable to conclude that Cl^- sits inside the cage whereas Br^- is in rapid exchange over the different faces of the tetrahedron.

More generally, understanding and prediction of the nature of the complexes as a function the partners and environment is a very challenging computational problem. Whereas X-ray data confirm that the Cs^+ complexes of **222** are of inclusion type, NMR data in solution show that, depending on the solvent and on the counterion, the complexes may be either "inclusive" or "exclusive" [66]. A more dramatic case is the "**Barrel**, 6H^+ "/ NO_3^- complex where NMR data in aqueous solution provide convincing evidence for formation of an inclusive cryptate of 1:1 stoichiometry, but where X-ray analysis of the crystal shows that the NO_3^- anions have *external* coordination with the ammonium sites of the cage [67]! We are presently investigating these problems using MD simulations.

CONCLUSION

From the few applications seen above, it appears that modelling the conformations and binding properties of macro(poly)cyclic molecules brings new or deeper insights into the recognition properties of these species. The chemists who designed molecular receptors [1-3] spent much time working with CPK models [29] and it might be hoped that computer graphics systems replace or assist that activity. Cram quoted "From the beginning, we used Corey-Pauling-Koltun molecular models, which served as a compass on an otherwise uncharted sea of synthesizable target complexes. We have spent hundreds of hours building CPK models of potential complexes and grading them for derisability as research targets." [6]. Such a modelling activity is presently far from being a standard computerisable procedure, first because of technical difficulties such as synchronous bond rotations keeping the (poly)cycles closed, and second because the inspired way the structure is made and manipulated is hardly codable. From different techniques used to built selected conformers of "supermolecules" with constraints and used to sample their conformational space in vacuo, dynamic views of the free receptors and of the concept of lock and key complementarity emerge. Attempts are made to evaluate conformational and complexation energies, but care must be taken about the significance of these numbers because of the theoretical approximations used to simulate the "gas phase" situation. In addition the relevance of gas phase data for solution properties can be questioned. Conformational analysis in solution should benefit from recent theoretical tools and increased computer powers. The potential of mean force for "effective" rotation barriers or intermolecular interactions in solution will be calculated on prototype fragments and incorporated in larger systems.

The prediction of structures in solution can be in principle achieved through MD simulations taking into account explicitly the solvent. However, from computer experiments in the gas phase, it is stressed that simulation times of a few hundreds of

ps will be too short to allow for significant conformational interconversions and that considerable computer means will be required. A simpler task is the analysis of solvation pattern for given conformers of the solute, a problem of particular interest given the ionophoric properties of many synthetic macrocycles, analogous the those of natural antibiotics [11]. In solid state structures, solvent molecules may be strongly coordinated to the macrocycle or to its complexed ion [11, 12] and have to be considered as part of the "supermolecule". Modelling solvation pattern can be in principle achieved through MC or MD techniques [59, 55]. We performed Monte Carlo simulations on **18-6** hydrated in its D_{3d}, C_i and C₁ conformations [57]. From these simulations, it was suggested that dissolution of the crystalline C_i form would lead to a conformational change, and that the D_{3d} conformer might be populated in solution. In addition, the water structure around **18-6** was *predicted* for the three conformers of the crown, and *confirmed later* by a X-ray study of a **18-6/PO₄H₃/6H₂O** complex [58]. Thus, even if the solvation involves dynamics aspects, it is clear that some of the solvent molecules are so firmly bonded that they are part of the structure. We are presently simulating solvation patterns of other prototype macrocyclic receptors.

ACKNOWLEDGEMENTS

G. W. thanks S. Boudon and P. Auffinger for their help in preparing the manuscript, and a NATO exchange grant 0478/82 with P. A. Kollman for supporting much of this work.

REFERENCES

- 1 C.J. Pedersen, *J. Am. Chem. Soc.*, **1967**, 89, 2495 and 7017; *Pure Appl. Chem.*, **1988**, 60, 450.
- 2 B. Dietrich, J.-M. Lehn, J.-P. Sauvage, *Tetrahedron Lett.*, **1969**, 2885 and 2889; J.-M. Lehn *Struct. Bond.*, **1973**, 16, 1; *Acc. Chem. Res.*, **1978**, 11, 49.
- 3 D.J. Cram, J. M. Cram, *Science*, **1974**, 183, 803.
- 4 Nobel Prize Citation; Nobel Festival, December 10, **1987**.
- 5 J.-M. Lehn "Supramolecular Chemistry- Scope and Perspectives. Molecules, Supermolecules, and Molecular Devices " (Nobel Lecture), *Angew. Chem. Int. Ed.*, **1983**, 22, 89.
- 6 D.J. Cram "The Design of Molecular Hosts, Guests, and their Complexes" (Nobel Lecture); *Science*, **1988**, 240, 760.
- 7 E. Fischer, *Ber.*, **1894**, 27, 2985.
- 8 L. Pauling, *Nature*, **1948**, 161, 707.
- 9 D.E. Koshland, *Adv. Enzym.*, **1960**, 22, 45 and *Ann. Rev. Biochim.*, **1968**, 37, 672.
- 10 W.P. Jencks, "Catalysis in Chemistry and Enzymology", Mc Graw Hill, New York 1969.
- 11 M. Dobler, "Ionophores and their Structures", Wiley Interscience Pub., New York, 1981.
- 12 F. Vögtle, W.M. Müller, W. H. Watson, *Topics Curr. Chem.*, **1984**, 125, 131.
- 13 G. Wipff, P. Kollman, *J. Am. Chem. Soc.*, **1982**, 104, 3249.
- 14 G. Wipff, P. Kollman, *Nouv. J. Chim.*, **1985**, 9, 457.
- 15 G. Wipff, J.-M. Wurtz, "The macrotricyclic receptors SC24, SC24.nH⁺ and their cryptates: molecular mechanics and molecular dynamics studies." *New J. Chem.*, **1989**, 000.

- 16 J.-M. Wurtz, G. Wipff, "Dynamic views of macrocyclic receptors: molecular dynamics simulations and normal modes analysis", in *Transport through membranes: Carriers, Channels, Pumps*, A. Pullman et al., ed. Kluwer Academic Publishers, 1988, page 1.
- 17 P. Auffinger, G. Wipff "High Temperature Annealed Molecular Dynamics Simulations as a tool for conformational sampling. Application to the 222 bicyclic cryptand", *J. Comput. Chem.*, 1989, 000.
- 18 J.-M. Wurtz, "Logiciels Infographiques de Modélisation Moléculaire. Application à l'étude d'un récepteur macrocyclique". Thèse de doctorat, Université Louis Pasteur, Strasbourg, 1988.
- 19 W.L. Jorgensen, *Acc. Chem. Res.*, 1989, 22, 184; P. A. Bash, U.C. Singh, R. Langridge, P.A. Kollman, *Science*, 1987, 236, 564; P. A. Bash, U.C. Singh, F. K. Brown, R. Langridge, P. A. Kollman, *Science*, 1987, 235, 574; J.A. Mc Cammon, *Science*, 1987, 238, 486; M. Mezei, D. L. Beveridge, *Ann. New York Acad. Sc.* 1980, 482, 1; W.F. van Gunsteren, *Prot. Eng.* 1988, 2, 5; M. Migliore, G. Corongiu, E. Clementi, *J. Chem. Phys.*, 1988, 88, 7766.
- 20 T.P. Lybrand, J.A. Mc Cammon, G. Wipff, *Proc. Nat. Acad. Sc. USA*, 1986, 83, 833. J.A. Mc Cammon, T.P. Lybrand, S.A. Allison, S. H. Northrup, *Biomolecular Stereodynamics III*, ed. R. H. Sarma, NY Adenine, page 227.
- 21 J. Van Eerden, S. Harkema, D. Feil, *J. Phys. Chem.*, 1988, 92, 5076
- 22 M.H. Mazar, J.A. Mc Cammon, T.P. Lybrand, *J. Am. Chem. Soc.*, 1989, 111, 55.
- 23 P.D.J. Grootenhuys, P.A. Kollman, *J. Am. Chem. Soc.*, 1989, 111, 4046.
- 24 P.D.J. Grootenhuys, P. A. Kollman, *J. Am. Chem. Soc.*, 1989, 111, 2152.
- 25 W.L. Jorgensen, S. Boudon, T.B. Nguyen, *J. Am. Chem. Soc.*, 1989, 111, 755.
- 26 U.C. Singh, P. Weiner, J. Caldwell, P. Kollman, AMBER_2.0, U.C. San Francisco, 1985.
- 27 S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta, Jr, P. Weiner, *J. Am. Chem. Soc.* 1984, 106, 765.
- 28 N. Gresh, P. Claverie, A. Pullman, *Theoret. Chim. Acta*, 1984, 66, 1; "Intermolecular interactions: from Diatomics to Biopolymers", B. Pullman Ed. New York: Wiley 1978. A E Howard, U. C. Singh, M. Billeter, P.A. Kollman, *J. Am. Chem. Soc.*, 1988, 110, 6984; A. Warshel, M. Levitt, *J. Mol. Biol.*, 1976, 103, 227; T. P. Lybrand, P.A. Kollman, *J. Chem. Phys.*, 1985, 83, 2923.
- 29 W.L. Koltun, *Biopolymers*, 1965, 3, 665.
- 30 S. Lifson, C.E. Felder, A. Shanzler, *J. Am. Chem. Soc.*, 1983, 105, 3866.
- 31 W.H.M. Uiterwijk, S. Harkema, B.W. van de Waal, F. Göbel, H.T.M. Nibbeling, *J. Chem. Soc. Perk. Trans 2*, 1983, 1843.
- 32 G.M. Crippen, "Distance geometry and conformational calculations", Research Studies Press Wiley, New York, 1981.
- 33 P.K. Weiner, S. Profeta, G. Wipff, T. Havel, I.D. Kuntz, R. Langridge, P. Kollman, *Tetrahedron*, 1983, 39, 1113.
- 34 J.-M. Wurtz, "Logiciels Infographiques de Modélisation Moléculaire. Application à l'étude d'un récepteur macrotricyclique". Thèse de doctorat, Université Louis Pasteur, Strasbourg, 1988.
- 35 "2D3D: a program to build 3D structures on the VAX+PS300 graphics system, based on the distance geometry approach.", J.-M. Wurtz, G. Wipff.
- 36 M. Billeter, T.F. Havel, K. Wüthrig, *J. Comput. Chem.*, 1987, 8, 132.
- 37 M. Billeter, A.E. Howard, I.D. Kuntz, P.A. Kollman, *J. Am. Chem. Soc.*, 1988, 110, 8385.
- 38 W. Braun, *Quater. Rev. Biophys.*, 1987, 19, 115.
- 39 M. Billeter, G. Wipff, unpublished results.
- 40 A. E. Howard, P.A. Kollman, *J. Med. Chem.*, 1988, 32, 1669.
- 41 R.J. Loncharich, E. Seward, S.B. Ferguson, F.K. Brown, F. Diederich, K.N. Houk, *J. Org. Chem.*, 1988, 53, 3479.
- 42 G. Chang, W.C. Guida W. C. Still, *J. Am. Chem. Soc.*, 1989, 111, 4379.
- 43 M. Saunders, *J. Comput. Chem.*, 1989, 10, 203; *J. Am. Chem. Soc.*, 1987, 109, 3150.
- 44 S.R. Wilson, W. Cui, J.W. Moskowitz, K.E. Schmidt, *Tetrahedron Lett.*, 1988, 29, 4373.
- 45 M. Nilges, A.M. Gronenborn, A.T. Brünger, G.M. Clore, *Prot. Eng.*, 1988, 2, 27; A.T. Brünger, G.M. Clore, A.M. Gronenborn, M. Karplus, *Prot. Eng.*, 1987, 1, 399; A. Di Nola, H.J.C. Berendsen, O. Edholm, *Macromolecules*, 1984, 17, 2044.

- 46 R. Perrin, C. Decoret, G. Bertholon, R. Lamartine, *Nouv. J. Chim.*, **1983**, 7, 263. M. J. Bovill, D.J. Chadwick, I.O. Sutherland, *J. Chem. Soc. Perk. Trans. 2*, **1980**, 1529. M. Badertscher, M. Welti, P. Portmann, E. Pretsch, *Topics Curr. Chem.*, **1986**, 136, 17.
- 47 R. Geue, S.H. Jacobson, R. Pizer, *J. Am. Chem. Soc.*, **1986**, 108, 1150.
- 48 J.D. Dunitz, P. Seiler, *Acta Cryst.*, **1974**, B30, 2739; E. Maverick, P. Seiler, B. Schweizer, J.D. Dunitz, *Acta Cryst.*, **1980**, B36, 615.
- 49 P.A. Kollman, G. Wipff, U. C. Singh, *J. Am. Chem. Soc.*, **1985**, 107, 2212.
- 50 J.-M. Wurtz G. Wipff "MDNM, an interactive program to display Molecular Dynamics and Normal Modes of Vibration on a Evans et Sutherland PS300 picture system."
- 51 M. Levitt, C. Sander, P.S. Stern, *J. Mol. Biol.*, **1985**, 181, 423; B. Brooks, M. Karplus, *Proc. Nat. Acad. Sc. USA*, **1985**, 82, 4995.
- 52 G. Wipff, D. Gehin, P. Kollman, *J. Am. Chem. Soc.*, **1989**, 111, 3011.
- 53 C. Sirlin, *Bull. Soc. Chim. Fr.* **1984**, 2,25. J.-M. Lehn, C. Sirlin, *J. Chem. Soc., Chem. Comm.*, **1978**, 949; *Nouv. J. Chim.*, **1987**, 1,1, 693.
- 54 J.F. Stoddart, *Annu Rep. Prog. Chem. Sect. B*, **1983**, 80, 353; I. Tabushi, *Tetrahedron*, **1984**, 40, 269; R. Breslow, *Adv. Enzymol.*, **1986**, 58, 1; I.O. Sutherland, *Chem. Soc. Rev.*, **1986**, 15, 63; F.M. Menger, *Topics Curr. Chem.*, **1986**, 136, 1; V.T. D'Souza, M.L. Bender, *Acc. Chem. Res.*, **1987**, 20, 146; J. Rebek, Jr, *Science (Washington DC)*, **1987**, 235, 1478.
- 55 C.A. Venanzi, *Int. J. Quant. Chem., Quant. Biol. Symp.*, **1986**, 12, 69.
- 56 D.H.J. Mackay, A.J. Cross, A.T. Hagler, "The role of energy minimisation in simulation strategies of biomolecular systems", Fassman ed., 1988.
- 57 M.L. Connolly, *J. Am. Chem. Soc.* **1985**, 107, 1118. B. Lee, F. M. Richards, *J. Mol. Biol.*, **1971**, 55, 379.
- 58 M.J. Heeg, *Acta Cryst.*, **1988**, C44, 2219.
- 59 E. Graf, J.-P. Kintzinger, J.-M. Lehn, J. Lemoigne, *J. Am. Chem. Soc.* **1982**, 104, 1672. E. Graf, J.-M. Lehn, *Helv. Chim. Acta*, **1981**, 64, 1040.
- 60 B. Owenson, R.D. Mc Elroy, A. Pohorille, *J. Am. Chem. Soc.*, **1988**, 110, 6992; B. Owenson, R.D. Mc Elroy, A. Pohorille, *J. Mol. Struct. (Theochem)*, **1988**, 179, 467.
- 61 W.L. Hase, M.-C. Richou, S.L. Mondro, *J. Phys. Chem.*, **1989**, 93, 539.
- 62 E. Graf, J.-M. Lehn, *J. Am. Chem. Soc.* **1976**, 98, 6403.
- 63 G. Ranghino, S. Romano, J.-M. Lehn, G. Wipff, *J. Am. Chem. Soc.* **1985**, 107, 7873.
- 64 E.H. Nordlander, J.H. Burns, *Inorg. Chim. Acta*, **1986**, 115, 31.
- 65 S.V. Hannongbua, B. Rode, *J. Chem. Soc. Farad. Trans. 2*, **1986**, 82, 1021; V. W. Riangpornvisuti, M.M. Probst, B.M. Rode, *Inorg. Chim. Acta*, **1987**, 134, 297.
- 66 E. Kauffmann, J.L. Dye, J.-M. Lehn, A.I. Popov, *J. Am. Chem. Soc.* **1980**, 102, 2274.
- 67 D. Heyer, J.-M. Lehn, *Tetrahedron. Lett.*, **1986**, 27, 5869.

DISCUSSION

GUILLOT - As far as small systems are considered, the evaluation of absolute free energies by various methods (perturbation methods, test particle method) is currently the subject of a debate. It seems now established that for neutral systems, these methods converge provided that the calculation is pushed far enough. On the contrary, for ionic solutions the perturbation approach presents serious difficulties related to the irreversibility of the path when charging progressively a molecule. In this context, do you think that your values of $\Delta\Delta g$ are meaningful, even if one can hope a cancellation of errors when the starting and the final states are not too different ?

WIPFF - This question refers to the calculation of *relative binding affinities in solution* for closely related substrates, bound to a same receptor (see references 19 to 25). I think indeed that our values for Cl^- versus Br^- binding to SC24,4H^+ are meaningful, since they are nearly identical to the experimental values (reference 20). There seems to be no particular problem if the mutation is performed keeping a constant charge, as we did. Because long-range interactions are calculated in practice with a cut-off distance, mutations with *variable* charges might lead to erroneous relative energies, but we kept a same charge along the mutations in water. Concerning the question of reversibility, we performed the mutation in the two directions ("forwards" and "backwards") to ensure that there is no hysteresis. Errors related to the energy representation, and particularly to the lack of polarization terms in the force field seem to cancel, since the agreement between relative energies calculated without polarization and the experimental values (e.g. relative solvation energies) is quite good (see references 19 to 25 and those cited therein).

BUCKINGHAM - Your potential model includes coulombic interactions and hydrogen bonds. Since the hydrogen bond may be considered to be electrostatic, how do you define a hydrogen bond in your calculations ?

WIPFF - Within the AMBER force field used, hydrogen bonding may be simply mimicked by electrostatic interactions (using a 1-6-12 potential) with appropriate choice of charges, or by using a 1-10-12 potential using slightly less polar charges. Both may be calibrated on H-bonded systems to reproduce gas phase enthalpies of complexation. However, as far as the dynamic behaviour is concerned, the distance dependence of the interactions leads to differences in dynamic properties with these

two representations. In the calculations reported here, we used a 1-6-12 potential with no particular choice of H-bond partners. Alternative calculations using the 1-10-12 representations for specific H-bonded atom pairs have also been tested (references 15 and 18). More specifically, these pairs involved N...+HN, O...+HN, N...HN, O...HN, H...OH contacts.

DEVILLERS - Among the 500 conformations you found by molecular dynamics on your system and after minimizations, did you find some you were able to consider as identical ?

WIPFF - We looked at this question only for experimentally known conformers. Indeed, when the High Temperature Annealed Simulations were performed on the free 222 cryptand, the experimental form II appeared twice. In the simulations on the 222/H⁺ cryptate, the **Ag** structure appeared 3 times, and in the simulations on the 222/K⁺ cryptate, the **K** conformer appeared 43 times.

PEPE - The crystal conformation of the cryptand is not representative of the solution conformations, may be this non correspondence is related to the charge distribution used in the coulombic energy calculation, as these molecule bear several heteroatoms.

WIPFF - The fact that the structures found experimentally in crystals of the 222 cryptates are not representative of the conformation observed *in solution* for the 222 cryptand is first due to complexation and environment effects : the conformations of cryptates (i.e. inclusion complexes) are induced by the complexed cation, and have converging orientations of the N,O binding sites. Second, because of dynamic effects in solution, the symmetrical structure observed *on the average* on the NMR time scale results from an equilibrium between non symmetrical forms. Our simulations on the *free* 222 show that structures with cavities have weak probabilities to be present in vacuo. Instead structures with no cavities, or with non-converging binding sites are found for free 222 in vacuo and there are good reasons to predict that they should be well hydrated and populated in solution.

As far as the electrostatic representation of the system is concerned, it is clear that cation complexation polarizes the C-O and N-O bonds, and that the dynamic behavior depends on the charges used in the calculations (see reference 16). Our set of charges used for free 222 describes reasonably the gas phase behaviour (see text).

Whether solvation would formally increase these charges (via polarization effects), or decrease them (via the dielectric constant) is not clear

DYMEK - When you monitored d_2 and d_1 as a function of time for one of your receptors, what was the rationale behind choosing those particular parameters ? Also what kind of information can you get (or what conclusion can be drawn) from the degree of correlation between the d_2 and d_1 trajectories ?

WIPFF - This question deals with the complex between cadaverine $+H_3N(CH_2)_nNH_3+$ ($n=5$) and its synthetic ditopic receptor (J.-M. Lehn *et al*, *J. Americ. Chem. Soc.* 1981, 103, 4266 ; *JCS Chem. Commun.* 1981, 833 and 1982, 557). We correlate the deformation of the cage with the motion of various substrates ($n=3$ to 7) and show that the correlation coefficient peaks for the substrate which is bound the most selectively ($n=5$). The choice of the distances which we correlated comes from the analysis of the low frequency normal modes of vibration in the complexes : the first mode indicates synchronous motions of the receptor and of the substrate (P. Auffinger, G. Wipff, unpublished results ; P. Auffinger, DEA 1987, Strasbourg).

MEHANI - It appears that you have successfully generated, using molecular dynamics, the solid state crystal structure conformers of cryptand [2.2.2] -potassium ion complex as a low energy conformer- If this structure has a mirror image conformer, then does your molecular dynamics studies generate both mirror image conformers equally ?

WIPFF - The **K** conformer of 222 is not generated from the simulations on free 222. It appears in the High Temperature Annealed MD Simulations only when K^+ is modelled in the cage. Although in principle both mirror image conformers should appear with equal probabilities in a purely random process, the 43 **K** forms generated have in fact the same chirality. This is because the K^+ cation prevents such interconversions on the time scale of 100 ps in vacuo. It is thus likely that in solution interconversion between mirror images of 222 occurs via decomplexation- complexation processes. It is also noteworthy that in the crystalline state, the NH_4^+ and K^+ cryptates of 222 display different enantiomeric forms of the cage.

CALCULATION OF THE FREE ENERGY BARRIER FOR THE ELECTRON TRANSFER REACTION
BETWEEN Fe^+ AND Fe^{2+} IN WATER

A. González-Lafont, J.M. Lluch, A. Oliva and J. Bertran

Departament de Química, Universitat Autònoma de Barcelona, 08193 Bellaterra
(Barcelona), Spain

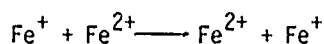
SUMMARY

For the $\text{Fe}^+/\text{Fe}^{2+}$ self-exchange reaction in solution it has been shown that the transition state is not reached starting from the reactants' region unless a conveniently designed sampling potential is employed. The free energy barrier has been calculated using two different techniques. The best value obtained using statistical perturbation theory is 176.5 kJ/mol.

INTRODUCTION

The classical treatment (refs. 1-2) of electron transfer reactions generally use an activated complex formalism. According to the Franck-Condon principle (ref. 3), internuclear and nuclear velocities do not change during an electron transition. Therefore, electron transfer must occur, within this classical frame, at the intersection region S^* of the diabatic potential hypersurfaces H_{pp} and H_{ss} , corresponding respectively to the precursor and successor complexes. This region S^* is reached by a suitable fluctuation in the nuclear configuration of the precursor complex. The electronic coupling matrix element between both diabatic states is assumed to be large enough so that the reactants are converted into products with unit probability in the intersection region, but small enough so that it can be neglected in calculating the amount of energy required to reach S^* (ref. 4).

In this paper we focus our attention on the outer-sphere self-exchange reaction:



in water. In this kind of processes, holding fixed the two metal ions, nuclear configurations are completely defined by solvent coordinates. To generate these configurations a statistical technique such as the Monte Carlo method is adequate. However, those fluctuations suitable for the electron transfer process have a very low probability and it is practically impossible to generate them upon starting from an equilibrium distribution of the solvent around the reactants (refs. 5-7).

Our purpose in this paper has been twofold. In first place, to show that even with a small cluster of 50 water molecules around both ions, S^* is not reached unless a conveniently designed sampling potential is employed (refs. 8-9).

On the other hand and in order to measure the probability that the system attains to the transition state region S^* , relative to the probability of being at the reactants' region, the free energy barrier is needed. In this way, we have used two different techniques to calculate the ΔF^\ddagger of the exchange reaction $\text{Fe}^+/\text{Fe}^{2+}$ in an aqueous solution including 125 water molecules.

METHODOLOGY

In this paper we have studied the self-exchange reaction between Fe^+ and Fe^{2+} ions which takes place in both a cluster of 50 water molecules and in a dilute aqueous solution including 125 water molecules. In the simulations the two ions are held fixed a distance of 7 Å. When simulating a dilute aqueous solution, the size of the box has been chosen in such a way that 125 water molecules give a density of about $1\text{gr}/\text{cm}^3$ and periodic boundary conditions under the minimum image convention have been used. To generate solvent configurations, the Monte Carlo method (ref. 10) within the Metropolis' algorithm (ref. 11) has been employed. In order to calculate the energies of the diabatic hypersurfaces Hpp and Hss, pairwise additive potential functions have been considered. We have used the MCY potential (ref. 12) for the water-water interaction, an ab initio analytical potential generated by us for the $\text{Fe}^+-\text{H}_2\text{O}$ interaction (ref. 13) and an empirical potential for the $\text{Fe}^{2+}-\text{H}_2\text{O}$ interaction (ref. 14). For all simulations a previous equilibration of the system followed by statistical analysis done over additional configurations, have been carried out.

To identify which of the Monte Carlo generated configurations belong to the intersection hypersurface S^* , the following computational procedure has been performed: For each one of the obtained solvent configurations of the precursor complex $(\text{Fe}^+/\text{Fe}^{2+})\text{aq}$, both the solvation energy of this complex and the solvation energy of the successor complex $(\text{Fe}^{2+}/\text{Fe}^+)\text{aq}$ obtained by keeping unchanged the water molecules coordinates but replacing the Fe^+ and Fe^{2+} ions by Fe^{2+} and Fe^+ , respectively, have been calculated. The difference between these two solvation energies gives the change (ΔE_{solv}) in solvation energy between Hpp and Hss for the frozen solvent configuration. Only those configurations with $\Delta E_{\text{solv}}=0$ are rigorously isoenergetic when the electron transfer takes place and therefore correspond to S^* .

RESULTS AND DISCUSSION

In first place we present the results obtained for the cluster of 50 water molecules in those simulations carried out computing the total energy by means of the Hpp potential. It corresponds to generate equilibrium nuclear configurations of the solvent around Fe^+ and Fe^{2+} in the precursor complex. For practical purposes, the wider criterion $|\Delta E_{\text{solv}}| \leq 5 \text{ kJ/mol}$ has been adopted in order to classify a given configuration as isoenergetic. To introduce the temperature effect we have performed simulations at 298 K, 350 K and 500 K. For each temperature we have generated 1,000,000 configurations in order to achieve equilibration and the corresponding analysis have been done over 1,000,000 additional configurations in each case. The energetic results for these statistical computations are presented in Table 1.

TABLE 1

Mean energy values^a of the cluster containing 50 water molecules at different temperatures.

T/K	298	350	500
E	-3471.8	-3327.2	-3062.3

^a In kJ/mol

Regarding the mean energy values obtained, we observe that when temperature is increased, regions of higher energy are populated. Nevertheless, no configuration corresponding to S^* has been identified. In this particular case, even if we use an exaggerated value of $|\Delta E_{\text{solv}}| \leq 20 \text{ kJ/mol}$, we do not find any isoenergetic configuration.

A stratagem that it is a very efficient method in order to generate configurations corresponding to the transition state consists in employing a sampling potential of the form (ref. 8):

$$U_{\lambda} = (1-\lambda) H_{\text{pp}} + \lambda H_{\text{ss}} \quad (1)$$

taking the parameter λ equal to $\frac{1}{2}$. In this way, we have carried out a simulation at 298 K on the cluster of 50 water molecules. 2,000,000 configurations have been needed to achieve equilibrium while the analysis have been done over 800,000 additional configurations. Using this potential, we have been able to identify a great number of configurations of the transition state when the above mentioned criterion of $|\Delta E_{\text{solv}}| \leq 5 \text{ kJ/mol}$ is adopted. The number of isoenergetic structures is about 66,400 (8.3%). This en-

semble of configurations presents a mean energy of -3319.7 kJ/mol which corresponds to the same energy range as the one populated by generating configurations with the Hpp potential at 350 K (see Table 1). However, no isoenergetic structure was found in that previous study while we have just mentioned the significant percentage obtained with this sampling technique.

The great amount of configurations classified as isoenergetic, enables us to study the structural and energetic characteristics of the transition state of our reaction. Computing the radial distribution functions of oxygen atoms around the two metal ions Fe^+ and Fe^{2+} , we have observed that both functions are nearly identical. The corresponding first peaks are located at 2.20 Å. This distance involves a mean solvation structure between those found for Fe^+ and Fe^{2+} ions in the precursor complex at equilibrium. These structural trends are in agreement with the fact that the sampling potential used tends to favour those symmetric configurations close to S^* . Finally, it has to be remarked that the transition state for this reaction cannot be associated to a unique well-defined structure. In fact, the configurations belonging to S^* present an important geometry dispersion as well as an energy dispersion which it is shown in Fig. 1.

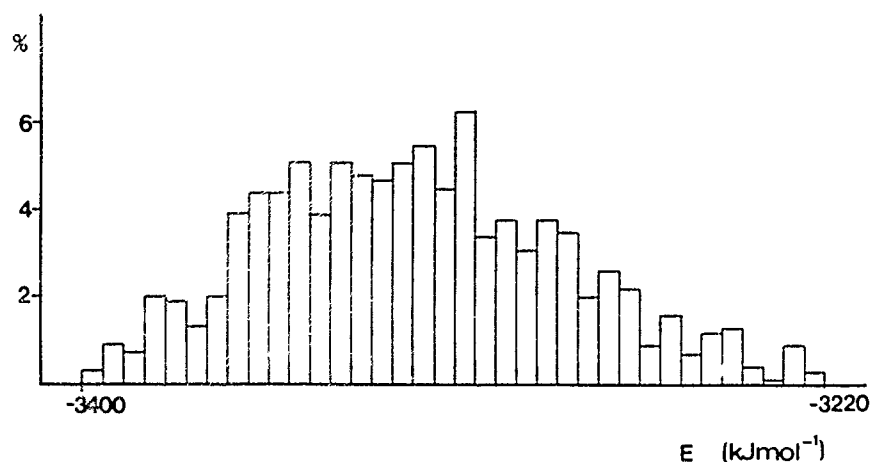


Fig. 1. Histogram of the percent distribution of isoenergetic configurations at 298 K against their energy.

Following the strategy explained above, we can analyze the nature of the transition state but we cannot evaluate its probability starting from the reactants' region. For this reason we have used two different methodologies applied to the dilute aqueous solution including 125 water molecules and periodic boundary conditions.

The first methodology is related with the parabolic behaviour of the free energy surface with respect to the solvent polarization coordinate assumed in

Marcus' theory (ref. 15). In this paper we have obtained the variation of ΔF with respect to the ΔE_{solv} parameter taken as the reaction coordinate. The extrapolation to $\Delta E_{\text{solv}}=0$ gives the activation free energy barrier. We have performed the simulations at 298 K generating 3,000,000 configurations for equilibrating the system followed by 2,000,000 structures for the analysis. While the configurations were being generated in the Monte Carlo process, we made a scanning of them as function of ΔE_{solv} . We have obtained configurations in a wide range of ΔE_{solv} values: From $\Delta E_{\text{solv}} = 891$ kJ/mol where 5 configurations are found ($2.5 \times 10^{-4}\%$) to $\Delta E_{\text{solv}} = 481$ kJ/mol with 21 configurations ($10^{-3}\%$). As it might be expected there was none of them with $\Delta E_{\text{solv}} = 0$. Collecting the generated configurations in intervals of 10 kJ/mol along the reaction coordinate ΔE_{solv} , we have identified the reactants' region with the most populated interval. This corresponds to the interval centered at 695.5 kJ/mol, 145790 configurations (7.29%) belonging to this region. If P stands for the number of configurations of each interval and P_{max} is the value associated with the reactants' interval, we define:

$$\Delta F = -RT \ln \frac{P}{P_{\text{max}}} \quad (2)$$

the factor $\exp(-\Delta F/RT)$ expresses the probability that the system will be in a given interval of the reaction coordinate ΔE_{solv} relative to the probability of being in the reactants' interval. We have fitted the following parabolic function by means of the least-squares method to the values of ΔF calculated with the above expression:

$$\Delta F = a(\Delta E_{\text{solv}})^2 + b\Delta E_{\text{solv}} + c \quad (3)$$

The parabolic behaviour of ΔF versus ΔE_{solv} is excellent as Marcus' theory had hypothesized (ref. 15) and as it has recently been shown by other authors in related systems (refs. 7-9,16). The mean absolute deviation obtained is only of 0.705 kJ/mol. Extrapolating to $\Delta E_{\text{solv}}=0$ we obtained the value of 207.3 kJ/mol for the free energy barrier ΔF^\ddagger . This value gives a relative probability of about 10^{-37} . This number reflects the fact that it was so difficult to reach the S^* region starting from reactants at equilibrium.

A more accurate result can be calculated by using statistical perturbation theory (ref. 17). This approach follows from eqn. (4) which expresses the free energy difference between systems 0 and n by an average of a function of their energy difference:

$$\Delta F = F_n - F_0 = -RT \ln \left\langle \exp \left(-\frac{(U_n - U_0)}{RT} \right) \right\rangle_0 \quad (4)$$

The average is for sampling based on system 0, so system \underline{n} is treated as a perturbation. In this paper we find the free energy associated with changing $U_0 = H_{pp}$ (unperturbed system 0) to the potential $U_{0.5} = 1/2 H_{pp} + 1/2 H_{ss}$ (perturbed system \underline{n}) that favours the appearance of isoenergetic structures. To avoid large perturbations, the coupling parameter of eqn. (1) has been used to smoothly transform the potential, therefore simulations have been run for $\lambda_i = 0.0, 0.1, 0.2, 0.3, 0.4$ and 0.5 in both directions, $\lambda_i \rightarrow \lambda_{i+1}$ and $\lambda_{i+1} \rightarrow \lambda_i$, except at the two end points. Double-wide sampling was used (ref. 18). In all, a number of six simulations has been performed. Each run has consisted of an equilibrium phase of 1,000,000 configurations followed by averaging over 1,000,000 additional configurations. The computed free energy changes are presented in Table 2.

TABLE 2

Computed free energy differences^a for the transformation from the unperturbed to the perturbed system.

λ_i	λ_j	ΔF	
		$i \rightarrow j$	$j \rightarrow i$
0.0	0.1	68.1	62.9
0.1	0.2	59.4	49.1
0.2	0.3	34.7	35.9
0.3	0.4	19.1	12.8
0.4	0.5	9.1	1.8
Total:		190.4	162.5

^a In kJ/mol

It is interesting to note that different values are obtained for the two independent runs, one in each direction of λ . However, this hysteresis is not very significant taking into account the large free energy barrier involved in the reaction. Anyway, to avoid systematic errors, the average value of 176.5 kJ/mol between those obtained in both directions is adopted

as the most valid for the free energy barrier of the $\text{Fe}^+/\text{Fe}^{2+}$ self-exchange reaction in solution.

Finally, it should be noticed that although the extrapolation method gives a somewhat exaggerated value in this case, it provides a reasonable approach to the value obtained with a more precise method.

REFERENCES

- 1 R.A. Marcus, J. Chem. Phys., 24 (1956) 966.
- 2 R.A. Marcus, J. Chem. Phys., 43 (1965) 679.
- 3 W.F. Libby, J. Phys. Chem., 56 (1952) 863.
- 4 B.S. Brunshawig, J. Logan, M.D. Newton and N. Sutin, J. Am. Chem. Soc., 102 (1980) 5798.
- 5 A. Warshel, J. Phys. Chem., 86 (1982) 2218.
- 6 A. González-Lafont, J.M. Lluch, A. Oliva and J. Bertran, J. Phys. Chem., 93 (1989) 4677.
- 7 R.A. Kuharski, J.S. Bader, D. Chandler, M. Sprik, M.L. Klein and R.W. Impey, J. Chem. Phys., 89 (1988) 3249.
- 8 J.-K. Hwang and A. Warshel, J. Am. Chem. Soc., 109 (1987) 715.
- 9 J.-K. Hwang, G. King, S. Creighton and A. Warshel, J. Am. Chem. Soc., 110 (1988) 5297.
- 10 J.P. Valleau and S.G. Wittington, Equilibrium Techniques, in: B.J. Berne (Ed.), Statistical Mechanics. Part A, Plenum Press, New York, 1977, chap. 4.
- 11 N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller and E. Teller, J. Chem. Phys., 21 (1953) 1087.
- 12 O. Matsuoka, E. Clementi and M. Yoshimine, J. Chem. Phys., 64 (1976) 1351.
- 13 A. González-Lafont, J.M. Lluch, A. Oliva and J. Bertran, Int. J. Quantum. Chem., 23 (1988) 77.
- 14 L.A. Curtiss, J. Woods Halley, J. Hautman and A. Rahman, J. Chem. Phys., 86 (1987) 2319.
- 15 R.A. Marcus and N. Sutin, Biochim. Biophys. Acta, 811 (1985) 265.
- 16 E.A. Carter and J.T. Hynes, J. Phys. Chem., 93 (1989) 2184.
- 17 R.W. Zwanzig, J. Chem. Phys., 22 (1954) 1420.
- 18 W.L. Jorgensen and C. Ravimohan, J. Chem. Phys., 83 (1985) 3050.

DESIGN OF NEW EFFICIENT RADICAL CLOCKS WITH THE HELP OF MOLECULAR MECHANICS

J.-M. MATTALIA, M. FATHALLAH, A. SAMAT, B. BLAIVE and M. CHANON

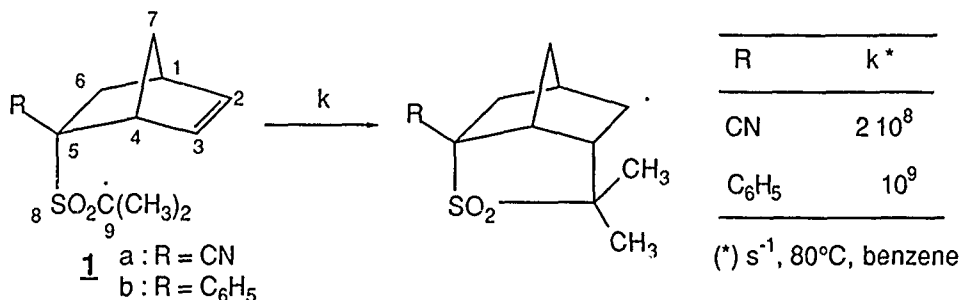
Faculté des Sciences S^t Jérôme. U.R.A. 126 CNRS. 13397 Marseille Cédex 13
(France)

SUMMARY

BLEMO, a molecular mechanics computer program operating with the MM2 parametrization, is used in order to design new kinds of radical traps involving intramolecular addition of the radical on a double bond. The 2-norbornenyle series is specially studied. Data show that the substitution on positions 5 and 7 with bulky groups and structures of the dimethanonaphthalene type lead to very short distances between reactive sites, and thus, probably to efficient radical clocks.

INTRODUCTION

The use of radical clocks is a very convenient approach to check for the presence of radical intermediates during organic reactions (ref.1). In particular, alkyl halide precursors of cyclisable radicals have been successfully applied to mechanistic studies (ref.2,3). The efficiency of this kind of radical clocks is governed by their cyclization rate (i.e., intramolecular rate of addition of the radical on a double bond) : the greater is the rate of rearrangement, the greater is the possibility to trap radical intermediate. We recently synthesized (ref.4-6) 5-(*endo*)-*i*-propylsulfonyl-5-(*exo*)-substituted-2-norbornene which leads to cyclization rates near 10^8 - 10^9 s⁻¹ for the corresponding C-centered α -sulfonyl radicals **1** (Scheme 1). In order to monitor the occurrence of short lived radical

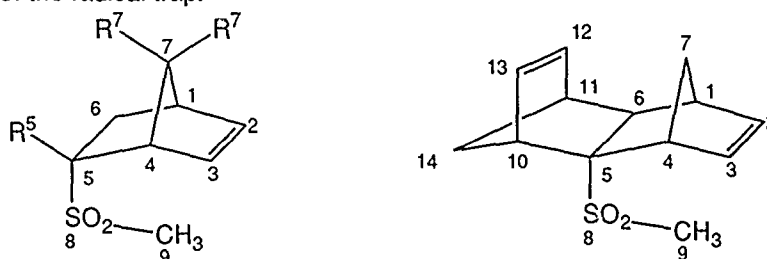


Scheme 1 Intramolecular cyclization of radical traps of the 5-(*endo*)-*i*-propylsulfonyl-5-(*exo*)-substituted-2-norbornene series (ref.5).

intermediates during reactions for which S.E.T. was postulated (in particular when solvent caged radical pairs are involved (ref.4,5,7)) we aim at designing and synthesising

free radical clocks of increasing efficiency higher rates of cyclization .

Our previous results led to the question : why does the substitution of a cyano group by a phenyl group in the 5-*exo*- position of 2-norbornene induce an increase in the cyclization rate of the radical **1** (scheme1) ? X-ray determination and force field calculations (ref.8) performed on both parent H compounds (ref. 6) suggest that this enhancement is mainly due to the structural constraints imposed by the phenyl group. In particular the modification of the valence angle $C_4-C_5-S_8$ (117.5° and 108.6° respectively for cyano and phenyl substituents), pushes the whole *endo*- substituent toward the double bond of the norbornenyl group, and puts closer the reactive sites, i.e. C_9 and C_3 . Taking into account these results, we decided to investigate, from the norbornenyl structure as basis, new substitutions able to modify geometrical parameters and so, to lead to an extended range of cyclization rates. The influence of the substitution on positions 5-*exo*, 7 and the interest of dimethanonaphtalene was investigated (scheme 2) in order to decrease the minimal distance between the reactive centres of the radical trap.



Scheme 2 : Norbornenyl structures studied with the help of molecular mechanics.

METHODOLOGY

The geometry of each molecule was optimized with BLEMO (ref. 9) a molecular mechanics computer program (ref. 8a) operating with the MM2 parametrization (ref. 10). With this parametrization, the energy calculated by BLEMO for a given geometry is the same as the one calculated by MM2, as it was verified for various test molecules. In BLEMO we decided to take into account all Van der Waals interactions at each step of the minimization, whereas in MM2 these interactions are restricted to the distances $r < r_{\max} = (12)^{1/2} \text{ \AA}$ in the intermediate calculations. Indeed, we preferred to calculate the energy in a strictly identical manner during the course of the minimization and in the final energy calculation at the optimum. It is interesting to note that this choice displaces very slightly the optimal geometry and the corresponding energy (about 0.1 kcal/mole). On the other hand, in BLEMO, the geometric variables for the optimization are internal coordinates (cartesian coordinates are taken in MM2). The starting geometry is built automatically, simplifying the entry of the molecule. In addition, the program gives the energy localization in the form of internal energies and interaction energies of predefined

fragments in the molecule.

After the optimization, the conformation of the lowest total energy is obtained. In order to know the minimal distance between the reactive sites (C_3 and C_9) corresponding to this structure, a rotation around the C_5-S_8 bond is performed leading the C_9 atom in the same plane as C_3 , C_5 and S_8 ($\widehat{C_3C_5S_8C_9} = 0^\circ$) without further minimization (fig. 1).

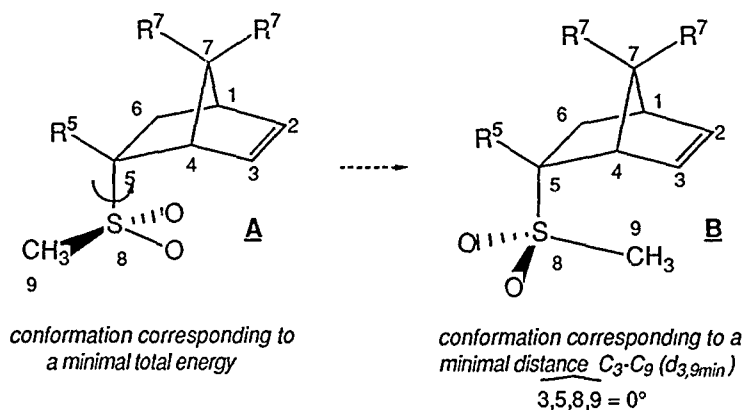


fig. 1 Example of rotation around the C_5-S_8 bond (from the conformation corresponding to the lowest total energy) leading to the minimal distance between C_3 and C_9 ($d_{3,9min}$). Depending upon R^5 the 9- CH_3 is more or less staggered with respect to the C_5-R^5 bond in **A** whereas it is clearly transoid in **B**.

RESULTS AND DISCUSSION

If the norbornene is substituted in the position 7 with alkyl groups (CH_3 , $i-C_3H_7$), phenyles or with spiro systems like $(-CH_2)_4$, $(-CH_2)_5$, the minimal distance available between C_3 and C_9 decreases when the valence angle 3,4,5 decreases (table 1). Bulky groups push on themselves and close the norbornene. In contrast, the substitution with a three links spiro $((-CH_2)_2)$, too constrained, opens the bicyclic system (angle 3,4,5 increases).

When the hydrogen on the position 5 is substituted by an alkyl group (CH_3 , C_2H_5), the methylsulfonyl group is pushed toward the double bond, the valence angle 4,5,8 decreases and as a consequence, the distance $d_{3,9min}$ decreases also (table 2). Intracyclic valence angles of the bicyclo are not noticeably changed. In the other hand, if the substitution on the position 5 with an ethyl group is combined with the addition of i -propyl groups on the methylene bridge, a very short distance between C_3 and C_9 (2.36\AA) may be obtained (table 3).

TABLE 1

Evolution of the minimal distance between C₃ and C₉ ($d_{3,9min}$) and of the angle 3,4,5 with the substitution on position 7 of the norbornenyle entity.

R ⁷	(CH ₂) ₂	H	(CH ₂) ₄	C ₆ H ₅	(CH ₂) ₅	Me	iPr
$d_{3,9min}(\text{\AA})$	2.62	2.59	2.57	2.57	2.55	2.55	2.51
$\widehat{3,4,5}(\text{°})$	110.1	109.8	108.9	109.1	108.4	108.4	107.3

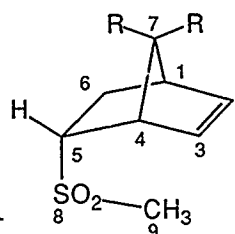


TABLE 2

Evolution of the minimal distance between C₃ and C₉ ($d_{3,9min}$) and of the angle 4,5,8 with the substitution on position 5-*exo* of the norbornenyle entity.

R ⁵	H	Me	Et
$d_{3,9min}(\text{\AA})$	2.59	2.51	2.48
$\widehat{4,5,8}(\text{°})$	112.8	111.9	110.1

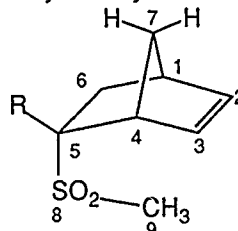


TABLE 3

geometrical parameters of the 5-ethyl-7-di-*i*-propyl norbornenyle entity.

$$d_{3,9min} = 2.36 \text{ \AA}$$

$$\widehat{3,4,5} = 107.8^\circ$$

$$\widehat{4,5,8} = 111.2^\circ$$

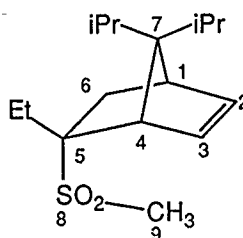
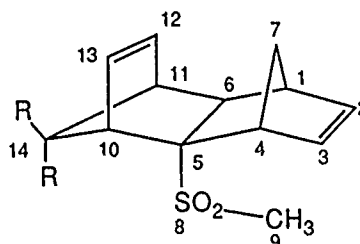


TABLE 4

Minimal distance between C₃ and C₉ ($d_{3,9min}$) and value of the angle 4,5,8 for dimethanonaphthalenes.

R ¹⁴	H	Me
$d_{3,9min}(\text{\AA})$	2.44	2.43
$\widehat{4,5,8}(\text{°})$	105.3	104.1



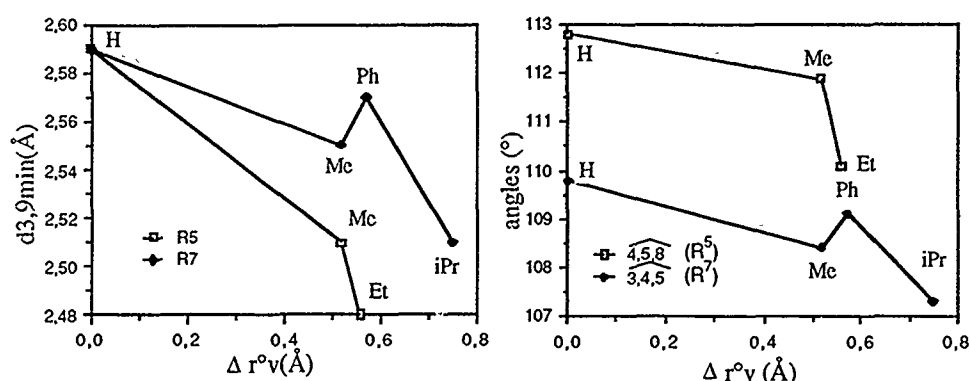


fig. 2 Diagram showing the relationship between the steric hindrance of the substituent in the 5- or 7- position ($\Delta r^{\circ}v$ is the corrected Van der Waals parameter = $r^{\circ}v(R) - r^{\circ}v(H)$ (ref.14)) and geometrical modifications of the norbornenyle entity (minimal distance between C_3 and C_9 ($d_{3,9min}$) and angles 4,5,8 (variation of R^5) or 3,4,5 (variation of R^7)).

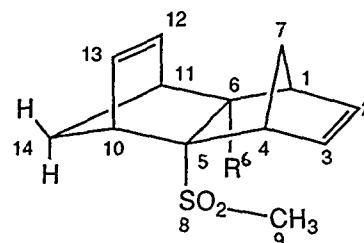
The influence of the steric hindrance of the 5- and 7- substituents on $d_{3,9min}$ and also on the exocyclic angle (4,5,8) or the intracyclic angle (3,4,5) is shown on fig. 2 where Van der Waals radii are taken as parameters for the size of the substituent.

Dimethanonaphthalene systems were also investigated, and in this case, the small value for the valence angle 4,5,8 is at the origine of short distances observed between the C-atom considered. If the substituent R^{14} (table 4) has only a little impact on $d_{3,9min}$, it might induce, from steric repulsions with the sulfonyl group, preferential conformations more favourable for the attempted cyclization. A methyl group placed on the position 6 leads to a geater variation of the distance between C_3 and C_9 (table 5). Two conformations of close energy must be considered. In the first, R^6 and S_8 are eclipsed (the dihedral angle $R^6,6,5,8 = 0^\circ$), and the resulting repulsion between the methyl and the methylsulfonyl groups brings about a decrease of the dihedral angle 6,4,5,8 and an increase of $d_{3,9min}$. In the second conformation R^6 and S_8 are slightly shifted, $\widehat{6,4,5,8}$ increases and $d_{3,9min}$ decreases.

TABLE 5

Influence of a 6-methyl group on the dimethanonaphthalene. Cf1 and Cf2 are related to the two conformations of lowest energy (E_{tot} is the total energy).

R^6	H	Me(Cf1)	Me(Cf2)
$d_{3,9\text{min}}(\text{\AA})$	2.44	2.53	2.40
$\widehat{6,4,5,8}(\text{^\circ})$	-119.6	-123.3	-120.0
$\widehat{R^6,6,5,8}(\text{^\circ})$	-2.3	0	-3.8
$E_{\text{tot}}(\text{kcal})$	75.8	78.8	79.2



CONCLUSION

This study shows that molecular mechanics calculations are an efficient tool for exploring the influence of the substitution on structural modifications on 5-(*endo*)alkyl sulfonyl-2-norbornenyl compounds. The results obtained point out the interest of bulky groups placed in positions 5 and 7 and of dimethanonaphthalenes in order to bring closer the carbon atoms 3 and 9 which are reactive sites in radical clocks of the norbornenyl series. Similar calculations were performed on molecules where the sulfone was replaced by a carbonyl function or a methylene group: the same general evolutions are found. This study therefore suggests the synthesis of new kinds of radical clocks of increased efficiency. If the introduction of bulky groups in the 7-position of the norbornene is synthetically difficult (ref. 11, 12), the substitution in the 5-position seems to be easier. We are performing experiments in this direction and in the dimethanonaphthalene series (ref.13).

REFERENCES

- 1 a) D. Griller and K.U. Ingold, *Acc. Chem. Res.*, 13 (1980) 317.
b) J.M. Surzur, in *Reactive Intermediates*, R.A., Abramovitch (Ed.), Vol. 2, Plenum, New York, 1982, pp. 121-295.
- 2 E.C. Ashby and J.N. Argyropoulos, *J. Org. Chem.*, 50 (1985) 3274 and references cited therein.
- 3 E.C. Ashby and T.N. Pham, *Tetrahedron Lett.*, 25 (1984) 4333.
- 4 B. Vacher, *Sciences Thesis*, 1987, Marseille, France.
- 5 B. Vacher, A. Samat and M. Chanon, *Tetrahedron Lett.*, 25 (1985) 5129.
- 6 B. Vacher, A. Samat, A. Allouche, A. Laknifli, A. Baldy and M. Chanon, *Tetrahedron*, 44 (1988) 2925.

- 7 C.Y. Meyers , in Topics in Organic Sulfur Chemistry, M. Tislev (Ed.), University Press, Ljubljana, Yugoslavia, 1978, pp. 207-260
- 8 a) J.B. Hendrickson, J. Am. Chem. Soc., 83 (1961) 4537.
b) N.L. Allinger, J. Am. Chem. Soc., 99 (1977) 8127.
c) E. Osawa, H. Musso, Topics in Stereochemistry, 13 (1982) 117.
- 9 C. Roussel, B. Blaive, R. Gallo, J. Metzger and J. Sandström, Org. Magnet. Reson. 14 (1980) 166.
- 10 N.L. Allinger and Y.H. Yuth, Molecular mechanics Program, QCPE 395, 1980.
- 11 R.W. Holder, J.P. Daub, W.E. Baker, R.H. Gilbert and N.A. Graf, J. Org. Chem., 47 (1982) 1445.
- 12 J.W. Wilt and S.Z. Amhed, J. Org. Chim., 44 (1979) 4000.
- 13 O. De Lucchi, N. Piccolrovazzi, G. Licini, G. Modena and G. Valle, Gazz. Chim. Italiana, 117 (1987) 401.
- 14 a) M. Charton, J Am. Chem. Soc., 91 (1969) 615.
b) A. Bondi, J. Phys. Chem., 68 (1964) 441.

DISCUSSION

DEVILLERS - Quel logiciel avez-vous utilisé pour faire ces calculs ?

M. FATHALLAH - C'est un programme de Mécanique Moléculaire écrit à l'ESIPSOI (Marseille) par B. Blaive et M. Fathallah portant le nom BLEMO. La paramétrisation est celle du programme MM2 d'Allinger complétée par une recherche personnelle.

DEVILLERS - Qu'offre de plus ce programme par rapport à MM2 ?

M. FATHALLAH -

. Le programme BLEMO est écrit en Fortran 77 mais il a des entrées- sorties très simples d'où une grande facilité d'utilisation.

. C'est un programme totalement automatisé ; si vous n'avez pas une géométrie de départ, il se charge d'en construire une.

. La méthode de minimisation est celle de relaxation ou "pas à pas".

. La minimisation se fait sur les coordonnées internes et non cartésiennes.

. BLEMO a été étendu à l'étude de l'atome de Fer et à d'autres hétéroatomes.

DEVILLERS - Comment peut-on se procurer ce programme ?

M. FATHALLAH - Il suffit d'écrire à B. Blaive ou M. Fathallah. De plus on a une version Mac qui sera disponible d'ici deux mois.

BRUNETIERE - Dans quelle mesure interviennent les énergies des différentes conformations ?

M. FATHALLAH - Les composés exocyano et exophényle 1 et 2 présentent entre leur conformation la plus stable et celle correspondant à la distance minimale des différences d'énergie relativement faibles (environ 5 kcal M⁻¹) et proches l'une de l'autre. Les populations des radicaux résidant à la distance minimale sont donc voisines pour 1 et 2.

PHASE BEHAVIOUR OF CYCLOHEXANE DERIVED FROM DIFFERENTIAL THERMAL
ANALYSIS AND COMPUTED BY MOLECULAR DYNAMICS

C. HOHEISEL¹; M. HARTMANN² and A. WÖRFLINGER²

¹Theoretische Chemie, Ruhr-Universität Bochum, Universitätsstraße
150, D-4630 Bochum, FRG

²Physikalische Chemie II, Ruhr-Universität Bochum, Universitäts-
straße 150, D-4630 Bochum, FRG

SUMMARY

Differential thermal analysis has been used to study the phase behaviours of cyclohexane, C_6H_{12} , and deuterated cyclohexane, C_6D_{12} , both in closed and open DTA-wells. The aim of the latter experiment was to investigate the possible influence of dissolved helium on the phase diagram. Both substances exhibit rotator and non-rotator phases as well as new high-pressure induced phase transitions.

A six-centre Lennard-Jones potential is proposed for cyclohexane and employed to molecular dynamics computations. Preliminary calculations concerning transport properties and the rotational behaviour predict the phase transition of liquid C_6H_{12} to the plastic phase realistically. However, the solid state points correspond rather to an amorphous phase. Therefore the potential was slightly adjusted in order to be compatible with a stable fcc structure.

INTRODUCTION

Cyclohexane is a well-known representative of orientationally disordered ("plastic") crystals. It exhibits a solid state transition at 186 K and atmospheric pressure, where the monoclinic crystal structure (solid II) changes to a high-temperature plastic phase (solid I) of fcc structure (ref. 1). Also metastable forms have been detected, depending on thermal treatment (ref. 2) and sample size (ref. 3). In the past new pressure-induced phase transitions have been found both in cyclohexane, C_6H_{12} , (ref. 4) and deuterated cyclohexane, C_6D_{12} (ref. 5). Whereas C_6H_{12} reveals only one new high pressure phase (denoted as solid III), two solid phases (III and IV) are observed in C_6D_{12} . The enthalpy changes for the low-temperature transitions solid II - III, solid II - IV, or solid IV - III are much less than for those to the plastic

phase solid I. Therefore we suppose that the high-pressure forms III and IV are not plastic phases.

Until now there exist no spectroscopic studies at high pressures that enable us to elucidate the crystal structure and molecular dynamics of the new high-pressure phases, apart from a recent neutron scattering experiment (ref. 6). On the other hand, computer simulations are increasingly employed to analyze crystal structure transformations as well as the features of rotator phases (ref. 7). Therefore we have performed molecular dynamics (MD) calculations in order to get more insight into the dynamic properties of the various phases of cyclohexane. Preliminary results (ref. 8) are encouraging and suggest further calculations.

EXPERIMENTAL

Differential thermal analysis (DTA) has been used to study the phase behaviours of C_6H_{12} and C_6D_{12} . In general transition temperatures are determined from DTA heating runs at a rate of 2 K/min. Details of the high-pressure set-up have been described previously (refs. 4,9). Pressure is generated by compressed argon, nitrogen or helium. The substance under investigation is enclosed in a small container of indium, the walls of which transmit the pressure but prevent the compressed gas from being dissolved in the substance. It is well known that solutions of compressed gases in organic compounds at high pressures may change the phase behaviour (ref. 10) and conformational equilibria (ref. 11). Because the neutron scattering study mentioned in the Introduction (ref. 6) has been performed under helium pressure, we reinvestigated the phase situation of C_6H_{12} and C_6D_{12} , using open DTA wells and helium as a pressure transmitting medium.

Within the limits of experimental error (± 0.5 K) no changes of the transition temperatures were observed. Both for C_6H_{12} and C_6D_{12} the DTA peak due to the solid III - I transition appeared sometimes a bit broadened, but an additional transition peak could not be resolved. However, a splitting of the solid solid transitions at lower temperatures (C_6H_{12} : II - III, C_6D_{12} : IV - III and very seldom also II - IV) occurred after several repeated runs. The longer the measuring time and the slower the heating rate the more frequently such splittings were observed. Examples are shown in Fig. 1. However, it should be mentioned that in the case of C_6H_{12} the splitting was only observed after repeated runs of

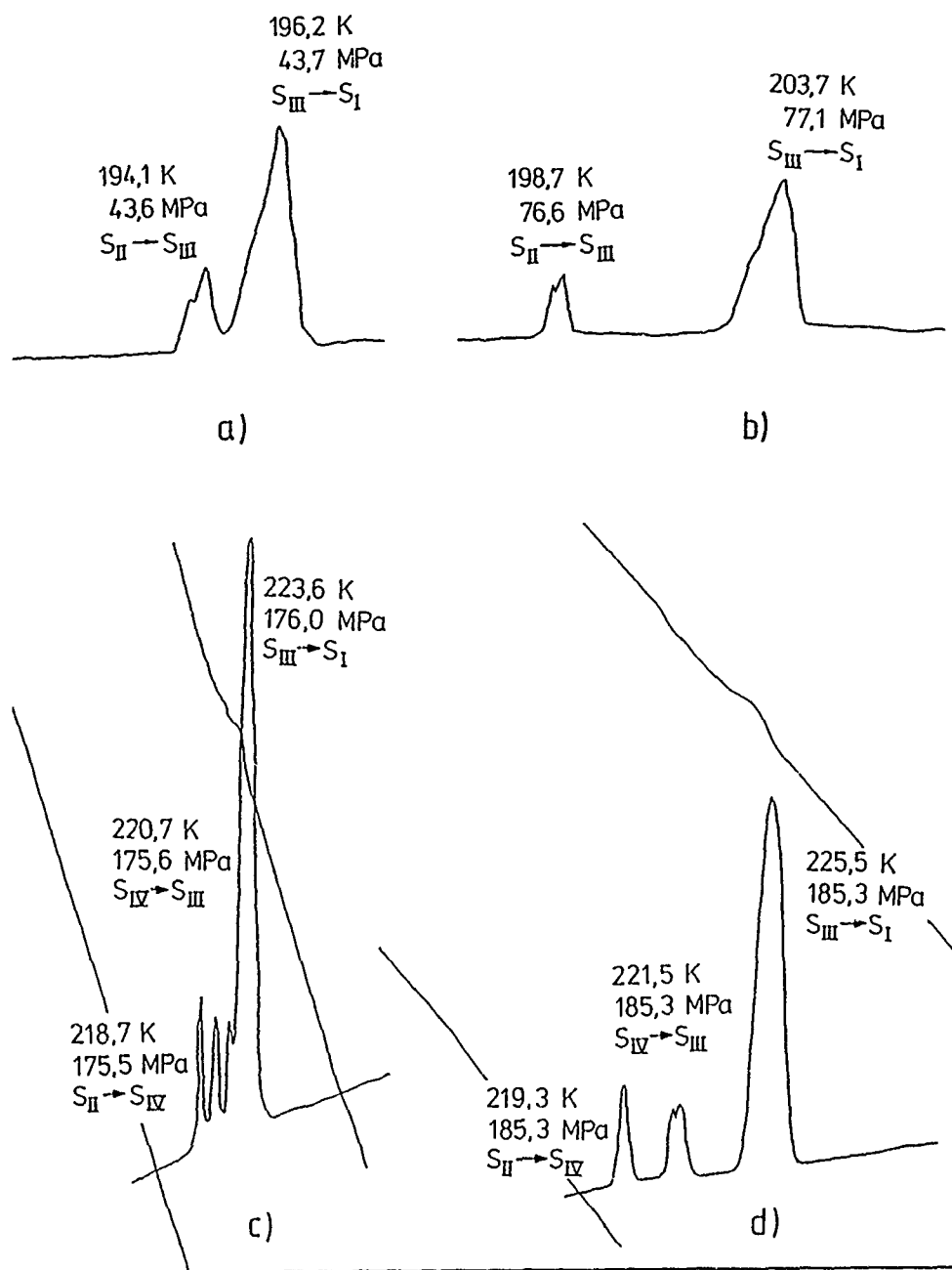


Fig. 1. DTA traces of C_6H_{12} (a,b) and C_6D_{12} (c,d) obtained in open wells under helium pressure on heating.

C_6H_{12} : traces after 30 cycles, rate 0.9 (a) and 0.7 K/min (b),

C_6D_{12} : traces after 8 cycles, rate 2.5 (c) and 1 K/min (d).

several days. It is possible that such long runs could introduce impurities in the open DTA wells that favour the occurrence of metastable phases (ref. 12).

Whether this finding is an artefact or an indication of further solid phases is still open to question. The possible existence of a fifth solid phase has recently been proposed for C_6D_{12} in the context of high pressure neutron scattering studies (ref. 13).

MD CALCULATIONS

For the MD calculations the method of constrained dynamics was applied, using the Stoermer-Verlet algorithm for numerical integration, as described in previous papers (refs. 14,15). A six-centre Lennard-Jones (LJ) potential is used for C_6H_{12} , each CH_2 -group being represented by a single LJ centre. Although this is an approximation, particularly for closed-packed crystals, such a model potential has been proven to be successful in predicting the phase transition from the liquid state to the rotator phase (ref. 8). Certainly further refinements of the model potential are necessary in order to explain the complicated high-pressure phase behaviours.

First results of the computations have been reported (ref. 8), some technical details are summarized in Table 1. Using the potential mentioned above we found that the fcc lattice of our model C_6H_{12} is not stable over a period of about 100 ps. This is indicated by the plots of the radial pair correlation function, $g(r)$, of the centres of masses given in Fig. 2a. Apparently the structure is changed towards an amorphous solid during the period of the MD computation of about 100 ps.

This drawback has been overcome by a further slight adjustment of the σ -parameter of the LJ-potential of each atomic site. Indeed Fig. 2b shows that the shape of the $g(r)$ curve is maintained over a computation period of 100 ps. Enlargement of σ by about 3% has two positive effects: It stabilizes the fcc lattice and brings the first neighbour separations (6.03 \AA) between the centres of two molecules in better agreement with experiment (6.09 \AA , ref. 1).

On the other hand, the slight adjustment of σ has only small influence on the thermodynamic quantities and the transport coefficients as we have checked previously for the liquid model. This improved model potential will be used in further computations of the phase behaviours of cyclohexane, especially for the low-temperature phase transitions.

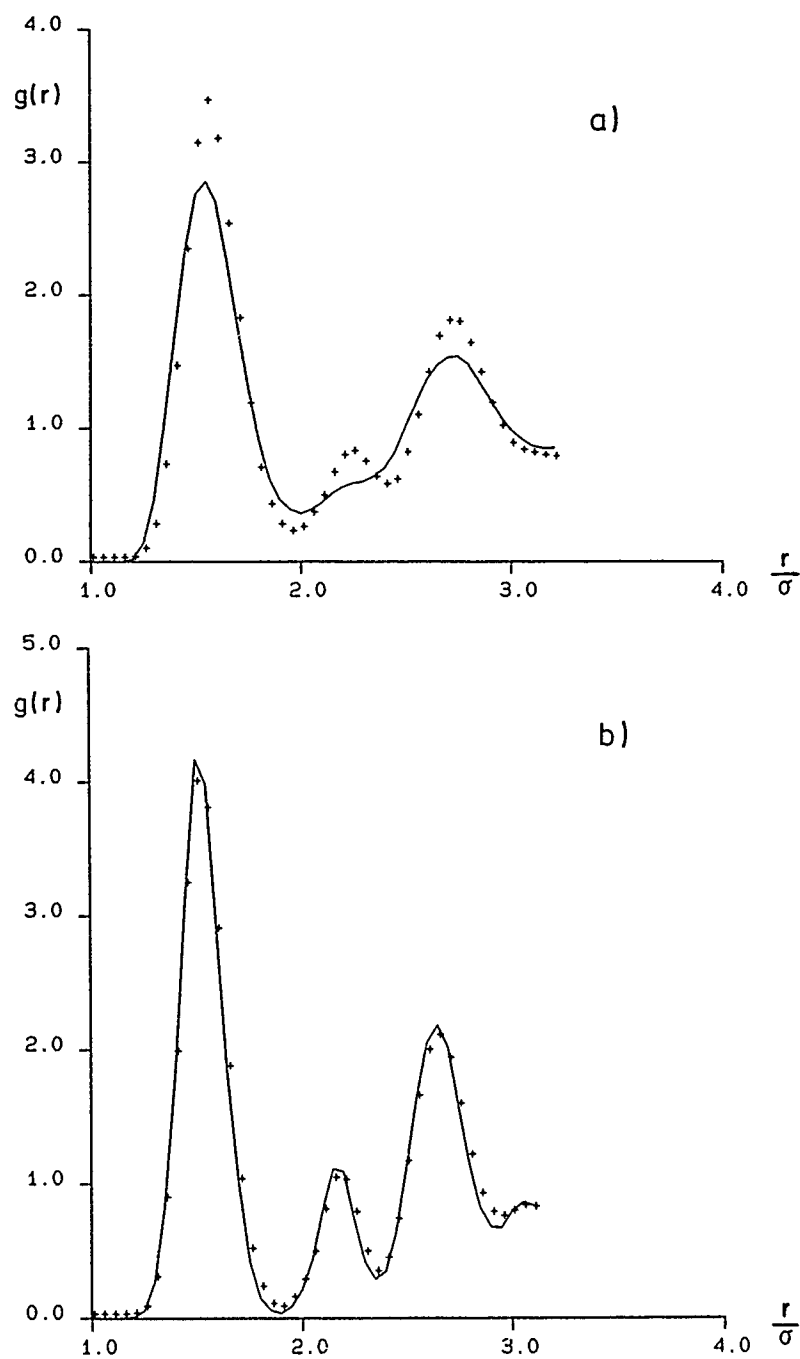


Fig. 2. Radial pair correlation function, $g(r)$, of the centres of masses (C_6H_{12}), $T = 198$ K, density = 0.8755 g/cm³; +++: after 7000 time steps, —: after 21000 time steps. (a) $\sigma = 3.86$ Å, (b) $\sigma = 3.97$ Å.

TABLE 1

Technical details of the MD computation

Time step	0.005 ps
Number of integration steps	7000 with N=108 for the single-particle properties, 80000 with N=32 for collective properties N = number of molecules
Structure of the C ₆ H ₁₂ model	chair form; separation between nearest LJ centres: 1.78 Å
LJ parameters	$\epsilon/k = 78 \text{ K}; \quad \sigma = 3.97 \text{ Å}$ $1 \text{ Å} = 10^{-10} \text{ m}$

REFERENCES

- 1 R. Kahn, R. Fourme, D. André and M. Renaud, *Acta Cryst. B* 29 (1973) 131-138.
- 2 M. Renaud and R. Fourme, *J. Chim. Phys.* 63 (1966) 27-32.
- 3 J.P. Dumas, *J. Phys. C: Solid State Phys.* 9 (1976) L143-146;
C. R. Acad. Sc. Paris, t. 284, Série C (1977) 817-820.
- 4 A. Würflinger, *Ber. Bunsenges. Phys. Chem.* 79 (1975) 1195-1201.
- 5 L. Schulte and A. Würflinger, *J. Chem. Thermodynamics* 19 (1987) 363-368.
- 6 J. Mayer, J. Chrusciel, S. Habrylo, K. Holderna, I. Natkaniec, M. Hartmann, A. Würflinger, S. Urban, W. Zajac, Poster presentation at the AIRAPT conference, Paderborn, FRG, 1989.
- 7 G. Ciccotti and W.G. Hoover (eds.), *Proc. Int. School of Physics: Molecular-Dynamics Simulation of Statistical-Mechanical Systems*, North-Holland, Amsterdam, 1986.
- 8 C. Hoheisel and A. Würflinger, *J. Chem. Phys.* 91(1) (1989) 473-476.
- 9 N. Pingel, U. Poser and A. Würflinger, *J. Chem. Soc. Faraday Trans. I*, 80 (1984) 3221-3231.
- 10 A. Würflinger and G.M. Schneider, *Ber. Bunsenges. Phys. Chem.* 77 (1973) 121-128.
- 11 M. Maissara and J. Devaure, *Fluid Phase Equilibria* 22 (1985) 231-329.
- 12 G.N. Zhizhin, E.L. Terpugov, M.A. Moskaleva, N.I. Bagdanskis, E.I. Balabanov and A.I. Vasilev, *Soviet Phys. Solid State* 14 (1973) 3028-3031.
- 13 S. Urban, *Inst. Phys., Jagellonian University, Cracow, Poland*, private communication.
- 14 C. Hoheisel and R. Vogelsang, *Computer Physics Reports* 8(1) (1988) 1-69; M. Schoen, C. Hoheisel and O. Beyer, *Mol. Phys.* 58 (1986) 699-709.
- 15 C. Hoheisel, *J. Chem. Phys.* 89(5) (1988) 3195-3202;
J. Chem. Phys. 89(12) (1988) 7457-7461.

THE Na^+ ION SOLVATION IN WATER-HEXAMETHYLPHOSPHORICTRIAMIDE MIXTURE:
MOLECULAR DYNAMICS SIMULATION.

I.I.VAISMAN, M.G.KISELEV, Y.P.PUHOVSKI, and Y.M.KESSLER

Institute of Non-Aqueous Solutions Chemistry of the USSR Academy of Sciences,
Akademicheskaya Street 1, 153045 Ivanovo (USSR)

SUMMARY

The molecular dynamics simulation technique has been applied to study the solvation of the Na^+ ion in binary solvent water-hexamethylphosphorictriamide (HMPT). Simulations with five different initial positions of ion were performed. Both time-averaged and time-dependent properties of water molecules were calculated in two regions of the MD box related to the polar and non-polar sites of the HMPT molecule and in the first and second solvation shells of the ion separately. Results are reported for various solute-solvent and solvent-solvent atom-atom radial distributions, hydrogen-bonded network structure, IR spectra and mutual arrangement of HMPT and cation. Preferential solvation of the sodium ion by HMPT was observed in one from five cases.

INTRODUCTION

Investigation of ionic solvation phenomena in binary mixed solvents at the microscopic level is essential for the understanding of physical properties of electrolyte solutions and their behaviour in many chemical, technological and biological processes. Numerous experiments have been devoted to studying this problem and even some generalizations were obtained (e.g.ref.1) but microscopic features of the events are still poorly understood. One of the most powerful tools for insight into this question is computer simulation.

In this work molecular dynamics simulation technique has been applied to the system Na^+ -water-HMPT. The main reasons of our interest in this system are as follows. HMPT ($\text{PO}(\text{N}(\text{CH}_3)_2)_3$, fig.1) has 6 hydrophobic CH_3 groups surrounding P atom and a highly polar $\text{P}=\text{O}$ group. Gutmann's donor number $\text{DN}=38.8$ and only few amines have higher DN's (ref.2). The oxygen atom can form 2 strong H-bonds with water molecules (according to IR and NMR spectra and dipole measurements in CCl_4). So, HMPT-water mixtures are characterized by strong solvent-solvent interactions. HMPT is unique in combination of both hydrophobic and hydrophylic types of hydration with high (nearly spherical) symmetry of its molecules. These types of interactions and structural peculiarities are reflected in a number of physical quantities of the solution. Investigations of systems containing HMPT and water as a mixed solvent can help to solve the long standing

problem on a mutual dependence of both types of hydration.

HMPT has high solvating ability for cations and it can be expected that selective solvation of Na^+ ion can occur in a water-HMPT mixture. The fact that the sodium ion is preferentially solvated by HMPT was confirmed by NMR studies (ref.3) and we made attempt to verify this in computer simulations.

Earlier we applied molecular dynamics method for water-HMPT solution (ref.4) and for Na^+ ion in water. Our results were in good agreement with experiment.

DETAILS OF THE SIMULATION

Molecular dynamics

The system considered in this paper consists of an ion, one HMPT and 205 water molecules in a cubic box with periodic boundary conditions. Its site length of 1.86nm corresponds to an experimental density of the H_2O -HMPT solution with molar fraction of HMPT $x=1/206 \approx 0.0049$ at 300K. The equations of motion were integrated using a leap-frog algorithm (ref.5) with a time step of 2fs. The SHAKE procedure (ref.6) was used to constraint the internal geometry of the HMPT and water molecules. The temperature was kept constant by coupling to an external heat bath (ref.7) with a time constant of 0.1ps.

HMPT molecule was inserted in the box contained the preliminary equilibrated system of 216 water molecules. The center of the HMPT molecule coincided with the center of the box, the initial coordinates of atoms of the HMPT molecule were taken from X-ray data. Water molecules having a distance less than 0.2nm from any atom of the HMPT were deleted. After insertion of the solute molecule an equilibration period of 4ps was applied.

Five molecular dynamics simulations labelled as A,B,C,D and E (Fig.1) of system sodium-water-HMPT were performed. (S is a mark for MD run of a water-HMPT system from ref.4, Na^+ - H_2O simulation marked as I). For each of five runs we choose a water molecule near either HMPT atom and replace it by a sodium ion. Initial distances between the Na^+ ion and two atoms of HMPT molecule are listed in Table 1. Each 5-ps run was executed after an equilibration procedure (usually 4-5ps). The length of a simulation limited by type of computer used, was chosen to give at least qualitative results on properties with short characteristic time.

Model parameters

The model potential for the water intermolecular interactions used in the simulation is the SPC model developed by Berendsen et al.(ref.8). A HMPT-HMPT intermolecular potential also of 12-6-1 type was defined (ref.4) in similar terms. The Lennard-Jones force constants were calculated using the energy

TABLE 1

Distances (in nm) between the Na^+ ion and HMPT atoms^a.

MD run	Initial		Equilibrium ^b	
	r_{IOX}	r_{IPH}	r_{IOX}	r_{IPH}
A	0.50	0.63	0.23	0.37
B	0.74	0.87	0.58	0.72
C	0.40	0.46	0.90	0.98
D	0.52	0.52	0.95	0.95
E	0.69	0.55	1.13	0.99

^aIOX and IPH refer to ion-oxygen and ion-phosphorous separations

^bDistances averaged over equilibrium configurations

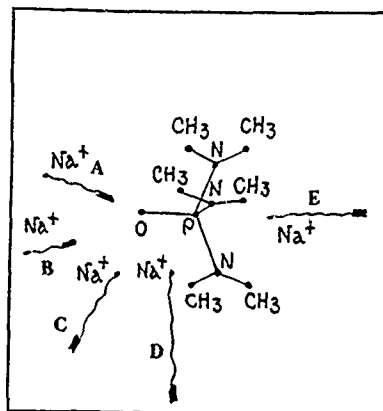


Fig. 1. Evolution of the ion position in the MD box during the simulation.

minimization of the HMPT dimer. Charges were chosen to give a satisfactory value for the dipole moment of HMPT. The methyl groups of HMPT were treated as united atoms. Potential used for ionic interactions was taken from ref.9. Interaction parameters of Na^+ ion with other atoms and those of water-HMPT interactions were obtained using combination rules. A cutoff radius of 0.9nm was applied for all interactions. "Shifted force" modification (ref.10) of potential has been employed to avoid truncation errors.

SOLVENT STRUCTURE

The complex character of the system leads to some differences in water distribution in either part of the MD box. Water-water radial distributions were calculated separately near the polar and hydrophobic sites of HMPT and in the first and second solvation shells of the sodium ion.

In our MD simulations we found three types of relatively stable positions of the sodium ion with regard to the HMPT molecule. Averaged distances between the Na^+ ion and the HMPT molecule in runs A and B correspond to positions of the first and second minima of the sodium-HMPT mean force potential. In runs C, D and E the ion is situated far from the HMPT molecule.

In the equilibrium configuration A the Na^+ ion replaces one of two water molecules in the first solvation shell of the oxygen atom of the HMPT. In the water-HMPT mixture these two molecules form hydrogen bonds with the HMPT oxygen atom. Presence of the Na^+ ion leads to distortion or even to destruction of the hydrogen bond between the second water molecule and the HMPT oxygen atom. There are in all four water molecules in the first solvation shell of the sodium ion in A configuration. Their arrangement is energetically advantageous and hence

it is stable. But orientations of these water molecules are mostly determined by both the ion and the HMPT oxygen atom, and links between water molecules from the first and second solvation shells are weak, average number of hydrogen bonds per water molecule is only 0.75. Results of this run (Tables 1,2) are very similar to those obtained from MD simulation of a Na^+ ion bound to a dihydrogen phosphate ion in a small water cluster (ref.11).

In run C we observed a hexahydrated sodium ion with the water molecules in the octahedral arrangement practically identical to one found in the $\text{Na}^+ - \text{H}_2\text{O}$ solution simulation (run I). But influence of the HMPT molecule leads to some changes in the orientation of the water molecules in octahedron apexes and therefore to the deviations in hydrogen bonds angles and energies.

By inserting of the sodium ion in the hydrophobic region of HMPT hydration (run E) one can observe that the first peak in the water oxygen-oxygen RDF near polar group of HMPT (Fig.2) as well as one in the HMPT oxygen - water oxygen RDF (Fig.3) becomes weaker than in water-HMPT simulation (run S). It can be concluded that sodium ion exerts a structure-breaking influence on the water network near the hydrophylic group of the HMPT molecule. It is interesting, that sodium - water oxygen radial distributions are similar in both mixed solvent and water (Fig.4).

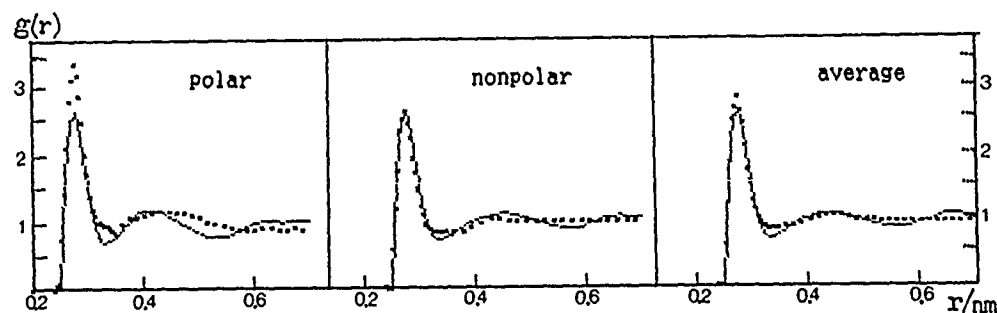


Fig. 2. Water oxygen-oxygen radial distribution functions near different sites of solute molecule and averaged over the whole box from runs S (dots) and E (full line).

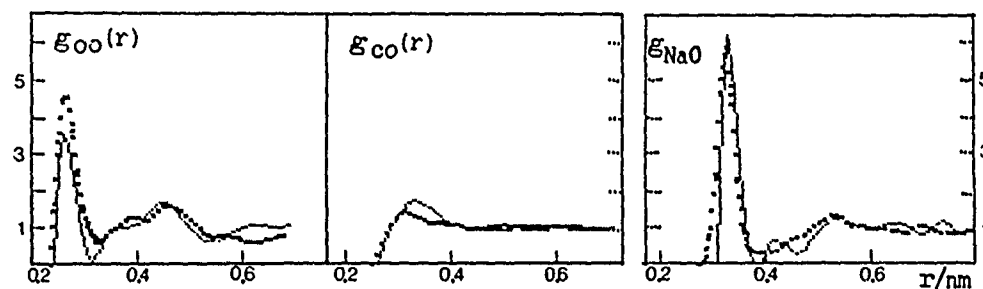


Fig. 3. Solute atom-water oxygen radial distribution functions (lines as in Fig.2)

Fig. 4. Ion-oxygen RDF from runs I (dots) and E (full).

An analysis of the hydrogen bonds gives an additional information about solvent structure. For defining the number of hydrogen bonds we considered only geometrical criteria ($r_{OH} < 0.2\text{nm}$ and angle $O-H-O > 140^\circ$). Orientational and energetic distributions of water hydrogen bonds around the HMPT molecule are displayed in Fig.5. Table 2 collects average numbers of hydrogen bonds per water molecule and some geometrical characteristics of water molecules arrangement in the Na^+ ion first solvation shell. Data from Table 2 and Fig.5 confirm the fact that in the presence of the ion water is more ordered in the non-polar region of HMPT hydration and less ordered in the polar one than without ion.

TABLE 2

Geometrical properties of the sodium ion water arrangement

Run	N	$\varphi_{\text{Na-O-H}}$		α	β	γ	N_{HB}	φ_{HB}
A	4.0	82	154	45	130	90	0.75	170
B	5.5	103	135	40	105	60	1.09	165
C	6.0	108	130	30	105	60	1.50	160

N=hydration number, $\varphi_{\text{Na-O-H}}$ =angles between Na-O and O-H vectors, α, β and γ = angles between Na-O vector and water molecules rotation vectors E_x , E_y and E_z respectively (x is coincide with direction of the dipole moment of water molecule, y is perpendicular to x in plane H-O-H and z is perpendicular to plane H-O-H, N_{HB} =average number of hydrogen bonds per water molecule in the first solvation shell, φ_{HB} =average angles of these hydrogen bonds.

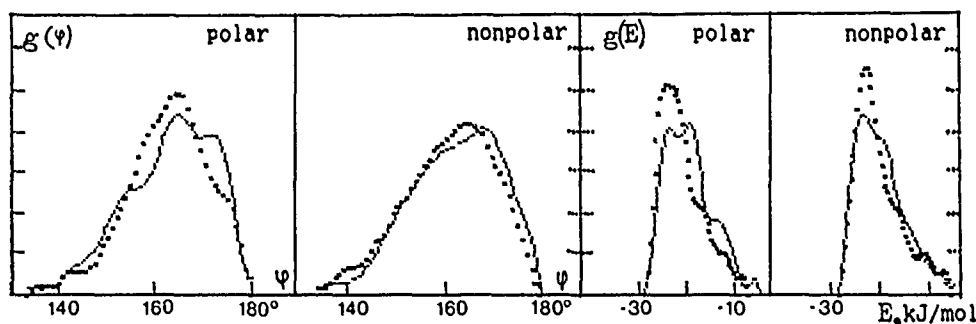


Fig. 5. Angular and energetical distributions of hydrogen bonds in water (lines as in Fig.2).

KINETIC PROPERTIES

To obtain further information about peculiarities of sodium solvation in mixed solvent, we have also calculated various auto- and cross-correlation

functions and their Fourier transforms. Some of them we discuss below.

Relaxation times of the dipole moments of water molecules in the first solvation shell of the Na^+ ion are very similar in runs A, B and C (Fig.6) and their value (2.5-2.6ps) is in good agreement with experimental one (ref.12) and with results of our MD simulation of water-HMPT solution (ref.4). But relaxation times of rotation of the unit vectors E_y and E_z are significantly different from those of dipole moment, so rotation of the shell water molecules is anisotropic and this anisotropy increases in row C, A, B. We suppose that this increase results from the dominant Coulomb interactions in the run A and from formation of common hydration shell of the sodium ion and the solute oxygen atom in the run B. Rotation of the shell water molecules in the runs D and E as well as in the run C is nearly isotropic, apparently because the ion is positioned relatively far from the HMPT in these cases.

The spectral densities of the hindered translations of the Na^+ ion, shell and bulk water molecules are shown in Fig.7. The diffusional component of Na^+ and shell water molecules motion is larger than for bulk water and for water in H_2O -HMPT solution. This difference in the diffusional motion can be related to the preferential solvation of Na^+ by HMPT.

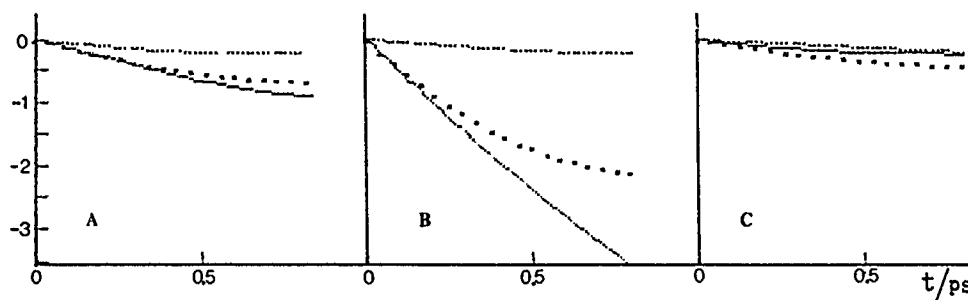


Fig. 6. Logarithm of the components of the shell water molecules rotational vectors autocorrelation functions: E_x (dashed), E_y (dots) and E_z (full). Axes are defined in Table 2 footnote.

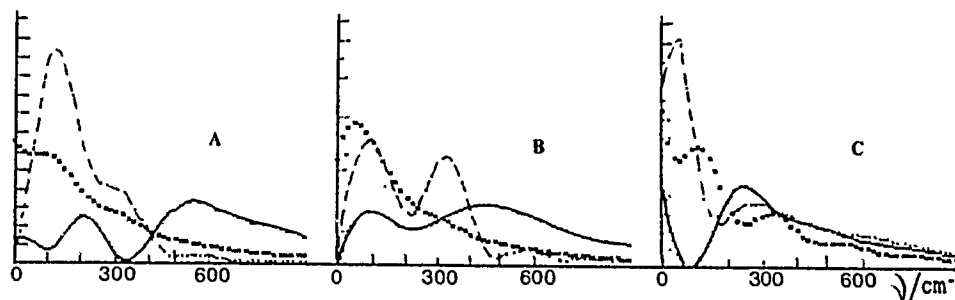


Fig. 7. Spectral densities (arbitrary units) calculated by Fourier transformation of the velocity autocorrelation functions of Na^+ (full line), shell water (dashed) and bulk water (dots).

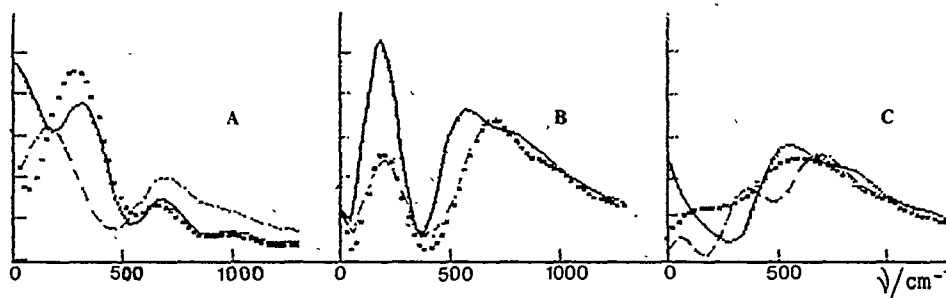


Fig. 8. Spectral densities (arbitrary units) calculated by Fourier transformation of the components of the shell water molecules angular velocities autocorrelation functions (see Fig.6).

The picture of preferential solvation is also supported by the investigations of librational motions of the shell water molecules (Fig.8). The normalized angular velocity autocorrelation functions have been Fourier transformed for the components in the three main axes of the water molecule (as described in the footnote of Table 2). The peaks corresponding to librations are shifted to higher frequencies for the runs A and B relative to the run C. Therefore the librational motions of the first solvation shell water molecules are influenced not only by the strong interactions with the ion but also by the interactions with the other molecules belonging to the and the second solvation shell, and the influence of the polar site of the HMPT prevail over the nearest water molecules influence.

It is impossible to investigate quantitatively the problem of ionic transport using short MD experiments such as this one. But some qualitative pictures elucidating the mechanism of ionic mobility can be found through direct observations as for example shown in Fig.9. This is one of the representative segments of the functions $r_{NaP}(t)$ and $N(t)$ from the run E. The value of N is defined simple as the number of the water molecules lying within the first solvation shell at this time step. One can see that the changes of the ionic velocity with respect to HMPT coincided with the extremas on the above curve in Fig.9 occur at the same time when either water molecule leaves or enters the first solvation shell. So, ionic motion is closely connected with local density fluctuations in the solvent. This result is in good agreement with our previous conclusions (ref.1).

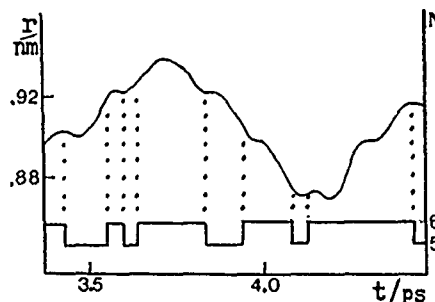


Fig. 9. Distance between Na^+ ion and HMPT phosphorus (above) and running hydration number of ion (below) .time.

CONCLUSIONS

Molecular dynamics simulations of a ternary system consisting of an ion, an organic molecule and water quantitatively shows in what way both the sodium ion and the HMPT molecule influence their water arrangement. This information can be very useful to answer to one of the fundamental questions of physical chemistry of multicomponent solutions: how are solute molecules of different types influencing the solvation of one another.

It was confirmed that the preferential solvation of Na^+ by HMPT can occur in the water-HMPT mixed solvent.

Patent correlation was found between diffusive displacements of the sodium ion and fluctuations of its coordination number. Hence we can study a link between the ionic mobility and the structure of the first solvation shell of this ion.

REFERENCES

- 1 Y.M.Kessler, R.S.Kumeev, I.I.Vaisman, R.B.Lyalina and R.H.Bratishko, Ionic mobilities at infinite dilution: structural aspects, *Ber.Bunsenges.Phys.Chem.*, 93(7) (1989) 770-782
- 2 V.Guttman, *The Donor-Acceptor Approach to Molecular Interactions*, Plenum Press, New York, 1978
- 3 M.Holz and A.Sacco, Molecular reorientation and Na^+ solvation in HMPT and HMPT-water mixtures, *Mol.Phys.*, 54(1) (1985) 149-160
- 4 M.G.Kiselev, I.I.Vaisman, Y.P.Puhovski and Y.M.Kessler, Thermodynamic properties of water-HMPT mixture from molecular dynamics, in: G.A.Krestov (Ed.), *Thermodynamics of Nonelectrolyte Solutions* (in Russ.), Ivanovo, 1989, pp.79-84
- 5 R.W.Hockney and J.W.Eastwood, *Computer Simulation Using Particles*, McGraw-Hill, New York, 1981
- 6 H.J.C.Berendsen and W.F.van Gunsteren, Molecular dynamics with constraints, in: J.W.Perram (Ed.), *The Physics of Superionic Conductors and Electrode Materials*, NATO ASI Ser.B92, Plenum, London, 1983, pp.221-240
- 7 H.J.C.Berendsen, J.P.M.Postma, W.F.van Gunsteren, A.Di Nola and J.R.Haak, Molecular dynamics with coupling to an external bath, *J.Phys.Chem.*, 81(8) (1984) 3684-3690
- 8 H.J.C.Berendsen, J.P.M.Postma, W.F.van Gunsteren and J.Hermans, Interaction models for water in relation to protein hydration, in: B.Pullman (Ed.), *Intermolecular Forces*, Reidel, Dordrecht, 1981, pp.331-342
- 9 J.Chandrasekhar, D.C.Spellmeyer and W.L.Jorgensen, Energy component analysis for dilute aqueous solutions of Li^+ , Na^+ , F^- and Cl^- ions, *J.Am.Chem.Soc.*, 106(4) (1984) 903-910
- 10 W.B.Streett, D.J.Tildesley and G.Saville, Multiple time step methods and an improved potential function for molecular dynamics simulations of molecular liquids, in: P.Lykos (Ed.), *Computer Modelling of Matter*, ACS Symp.Ser.v.86 American Chemical Society, Washington, 1978, pp.144-158
- 11 W.K.Lee and E.W.Prohofsky, A molecular dynamics study of the solvation of a Na^+ ion bound to PO_4H_2^- ion, *Chem.Phys.Lett.*, 85(1) (1982) 98-102
- 12 Y.M.Kessler, A.N.Petrov, I.I.Vaisman, V.V.Goncharov, G.A.Alper, Cation influence on the HMPT hydration shell, *Zh.Fiz.Khim.*, in press

SIMULATION OF PHASE TRANSITIONS IN ORGANIC CRYSTALS - THE GAMMA-ALPHA PHASE
TRANSITION OF P-DICHLOROBENZENE

J. ROYER*, F. BAYARD, C. DECORET and M. SKALLI

UA 805 C.N.R.S., BAT 305 and 308*, 43 Bd du 11 Novembre 1918, 69622
VILLEURBANNE CEDEX, FRANCE

SUMMARY

The different mechanisms concerning the phase transitions for molecular crystals are reviewed. In order to gain more insight in these mechanisms, we have calculated with an atom-atom potential method possible crystallographic paths for second and first order transitions.

1. INTRODUCTION

The ability of a substance to crystallize in different crystal forms gives rise to polymorphism. A variety of solids transform from one crystal structure to another and the subject of phase transitions has attracted special interest in recent years because it serves in other branches of science, as crystallography, geology, metallurgy, pharmacy etc ... (ref.1). We shall focus here on organic molecules where polymorphism is fairly common. In organic crystal mainly Van der Waals forces and hydrogen bonds occur. These type of solids are called molecular crystals. A transformation is designated as enantiotropic if the change from one polymorphic modification to another can be reversed in the solid state as the temperature or pressure is varied. If a change between two forms is irreversible in the solid state the forms are monotropic to each other. If the higher melting form has the lower heat of fusion the two forms are usually enantiotropic otherwise they are monotropic.

The transition of a material from one crystalline phase to another involves displacements of the molecules. These dynamic properties have currently been studied by a wide variety of experimental methods such as crystallography, N.M.R., second moment measurement, neutron scattering and far infra-red and Raman spectroscopies. The classification of phase transitions into various types has been prompted by the expectation that classification would facilitate an understanding of the mechanism of the microscopic nature of the phase transitions. The various classifications have been discussed by several authors (ref.2). It is well known that there are two kinds of phase transitions: the first kind called first order, in which energy, volume and crystal structure change discontinuously; the second kind called second order

in which energy and volume change continuously but the temperature derivatives of these quantities have singularities. Mechanistically, phase transformations classified as second order are generally associated with cooperative mechanisms and displacive transitions. The mechanisms of such phase transitions can be explained in terms of "soft" modes of lattice vibration. The agent of cooperative molecular reorientation is a strongly anharmonic lattice vibration. When the structure of both phases are known, it is possible to deduce the mechanism by finding a suitable soft mode and to devise an experimental method for detecting the softening as the transition point is reached. However only a small number of phase transitions have a soft mode behaviour. Chloranil (ref.3), malonitrile (ref.4), N-nitrodimethylamine (ref.5) exhibit soft mode behaviours confirmed by experimental measurements.

Mnyukh and coworkers (ref.6) deny the existence of any but first order phase transitions. They point out that differences in the phase transition behaviour can be the result of experimental inadequacies. In this view, a polymorphic transition occurs by way of growth of well shaped crystals in the parental crystal medium. The necessary condition for polymorphic transition is the presence in the crystal of lattice imperfections called nucleation centers of a definite quality and in a sufficient concentration. These lattice defects may be smaller than 1 micron and therefore may not always be visible under the optical microscope.

2. THE ATOM-ATOM POTENTIAL METHOD

A molecular crystal is characterized by the occurrence of molecules that keep their individual characteristics and properties. This is a consequence of the fact that the molecules in a crystal are bound together by forces that are very weak if compared to the intermolecular forces. As a consequence, the external and internal contributions to the crystal potential energy can be separated. The molecule-molecule interaction is split into the sum of different independent contributions, which are the attractive (dispersion) repulsive, electrostatic and polarization.

$$E = E_{\text{rep}} + E_{\text{disp}} + E_{\text{el}} + E_{\text{pol}}$$

All calculations involve the pairwise evolution of nonbonded interactions between atoms composing the molecules in the crystal. Among the most successful potential functions describing such interactions is the sixth power function which consists of a repulsive exponential part and an attractive portion

$$E = \sum_i \sum_j A \exp(-B R_{ij}) - C R_{ij}^{-6} + \sum_i \sum_j \frac{q_i q_j}{R_{ij}} - \frac{1}{2} \sum_i \sum_j \alpha_i (\epsilon_i)^2$$

where A, B and C are empirical parameters, R is the distance between atoms i and j, α^i is the mean polarizability attributed to atom i and ϵ_i is the electric field created at the atom i by all other molecules. The atom-atom potential method derived from static crystal properties may also be used to predict the lattice dynamics of organic crystals. It is an invaluable tool in the interpretation of infrared, Raman and neutron scattering experiments. The calculations have achieved success when applied to internal libration (ref.7), reorientations and orientational disorders (ref.8), dislocations (ref.9). Several reviews and books have summarized progress in this field in recent years (ref.10). These results encouraged us to explore whether potential energy calculation fits with the features of crystal polymorphic phase transformations. Our work is an attempt at such a fitting for both second order and first order phase transitions.

For example a possible crystallographic path for the polymorphic transformation of trans-trans (Z,Z) diacetamide into its trans-cis (E,Z) form has been calculated (ref.11). This transition is believed to be displacive. A plausible process for the phase transformation occurring between the two forms consists of two types of molecular movements. One is moving the centers of symmetry of pairs of molecules of the Z,Z form and the other movements are the rotation of the molecule as a whole and the internal rotation of the acetyl groups. The calculated barriers along this path are low enough to explain the ease with which diacetamide undergoes isomerization in the crystal. Now we will focus on a first order phase transition.

3. P-DICHLOROBENZENE PHASE TRANSITIONS

It is well known that the gamma \leftrightarrow alpha and the alpha \leftrightarrow beta transitions of p-dichlorobenzene occur with the nucleation growth processes. The gamma \leftrightarrow alpha transition has been studied by employing infrared and N.Q.R. spectroscopy as well as differential scanning calorimetry. Reynolds (ref.12) has postulated that a nucleus is produced by a series of fluctuations of large number of molecules over a large range. The size of the critical nucleus is about 1000 molecules. A nucleus reaching this size is likely to continue its growth to produce a macroscopic daughter crystal. It should be noted that Mnyukh and Reynolds admit the existence of a nucleation step. However Reynolds disagrees with Mnyukh and suggest that the nucleation step is often rate determining and the crystal growth requires mutual orientation of parent and daughter crystals whereas for Mnyukh the daughter crystal is

oriented randomly with respect to the parent crystal. For Reynolds molecular first order phase transitions are fundamentally martinsitic. The characteristics of a martinsitic transformation are that it is an heterogeneous transformation involving nucleation of a crystal and a subsequent rapid growth by a mechanism involving no long-range translational diffusion. These characteristics can be discussed by the possible epitaxy combinations due to the low symmetry in the phases and the presence of imperfections in the crystal retards the growth rates.

In this context we wish to consider in a first step the application of potential energy calculation to the modelling of nucleation for the gamma \leftrightarrow alpha phase transformation of p-dichlorobenzene. The gamma and alpha forms are reported from X-ray crystallographic determination (see Fig. 1) and G.L.Wheeler and S.D.Colson (ref. 13) suggest a possible way for the superposition of the two crystals alpha and gamma.

3.1 Superposition of gamma and alpha cells

According to Colson's ideas, we take for ab plane, the plane containing the molecule 1 (0,0,0) and the molecule 2 (0,1/2,1/2) obtained by an helical rotation. Now, cell parameters are changed and the gamma and alpha crystals can be described with the following doubled cells :

$$a=14.026 \text{ \AA} \quad b=6.021 \text{ \AA} \quad c=7.414 \text{ \AA} \quad \beta=102.72^\circ \quad P21/c \quad Z=4 \quad v=610.75 \text{ \AA}^3$$

$$a=14.664 \text{ \AA} \quad b=5.740 \text{ \AA} \quad c=7.850 \text{ \AA} \quad \beta=111.77^\circ \quad P21/a \quad Z=4 \quad v=613.62 \text{ \AA}^3$$

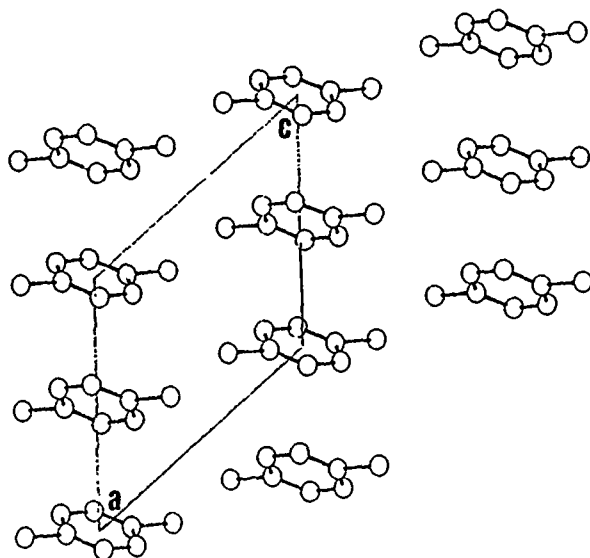
These doubled cells are necessary for calculating a crystallographic path (v.i.). The position of the centers of mass of the four molecules of the gamma and alpha forms are respectively (0,0,0), (0,1/2,1/2), (1/2,1,1/2), (1/2,1/2,0); (0, 1/2,0), (1/2,1,0), (0,1/2,1/2), (1/2,1,1/2).

Now the structure of the alpha form can be regarded as consisting of piles of molecules parallel to the c direction. These piles contain either molecules of type 1 (molecules 1 and 3 for instance for the pile (0,0,c)) or molecules of type 2 (molecules 2 and 4 for the pile (1/2,1/2,c)). It is to be noted that in one pile, molecules are parallel, molecules of type 2 being obtained by symmetry rotation from molecules of type 1. Furthermore, in each pile molecules are separated by 3.925 Å.

To rearrange in such piles, we may assume that molecules of the gamma crystal move in the bc planes along b. In such planes, molecules are arranged in piles in which their centers are deduced by translation of 7.414 Å along c, but each pile is shifted by 1/2 c (i.e 3.707 Å) so that they can be moved to imbricate one another in order to form alpha-like piles as shown in figure 1. In this assumption, molecules will build up new alpha-like piles at 1/4b, 3/4b and so forth. Accordingly, we displace the centers of mass of the four

Gamma form $P2_1/c$ $Z=2$

$a=8.624\text{\AA}$ $b=6.021\text{\AA}$ $c=7.414\text{\AA}$ $\beta=127.51^\circ$



Alpha form $P2_1/a$ $Z=2$

$a=14.664\text{\AA}$ $b=5.74\text{\AA}$ $c=3.925\text{\AA}$ $\beta=111.77^\circ$

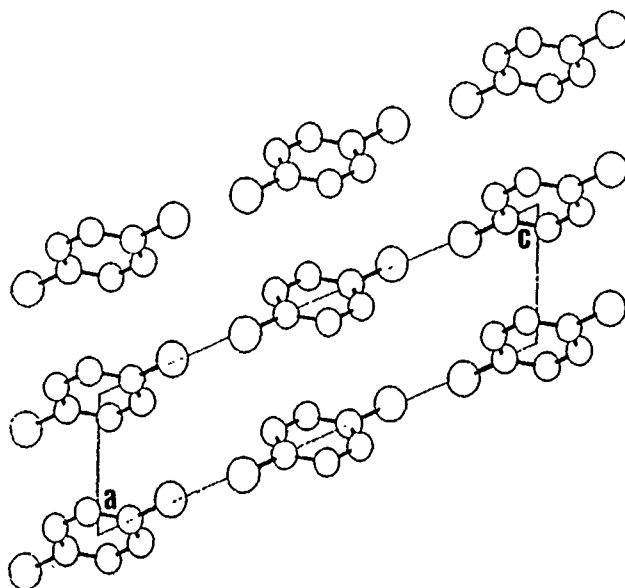


Fig. 1. Gamma and alpha forms of p-dichlorobenzene.

molecules of the alpha form and we have the new positions $(0, 1/4, 0)$, $(1/2, 3/4, 0)$, $(0, 1/4, 1/2)$, $(1/2, 3/4, 1/2)$. Finally we choose the superposition by coincidence of centers of mass on a and b directions. This superposition will permits us to displace and rotate molecules of the gamma phase unit cell, modifying also cells parameters in order to build an alpha phase, wich is described in the same referential as the gamma phase.

3.2 Superpositions of molecules.

The molecules 1 and 3 of the unit cell of the gamma form are normally deduced from one another by a translation $1/2$ along the a axis. To coincide the molecules 1 and 3 of the gamma form with the molecules 1 and 4 of the alpha form respectively, we have to move in opposite directions, the molecule 1 in the -b direction and the molecule 3 in the +b direction. The same is true for the molecule 2 (+b direction) and the molecule 4 (-b direction) of the gamma form to coincide with the molecules 3 and 2 of the alpha form respectively. These movements explain the need for doubling the gamma and alpha original cells.

To achieve a complete coincidence, three kinds of geometrical transformations are needed :

- displacement of the centers of mass,
- rotation of the Cl-Cl axis of the molecules in order to make coincide all chlorine atoms,
- rotation of the molecular planes around the Cl-Cl axis to make coincide all atoms.

The main features in this process is the weak rotations (Cl-Cl axis and molecular plane) for the molecule 1 of the gamma form to coincide with the molecule 1 of the alpha form and consequently for the molecule 4 of the gamma form to coincide with the molecule 2 of the alpha form. At the reverse, the molecules 2 and 3 of the gamma form have to perform large rotations to coincide with the molecules 3 and 4 of the alpha form.

We have now to consider how such displacements, in parallel with cell parameters transformations, can be combined to achieve the transformation of the gamma cell to the alpha cell.

We propose here a very simple way to perform a progressive transformation step by step. We prepared ten steps with regular $1/10$ th. modification of each geometrical transformations as described above (displacements and rotations), volume is also regularly increased of $1/10$ th for each step by suitable modifications of the cell parameters ($V = abc \sin(\beta)$). An atom-atom potential calculation with energy minimisation is then performed for each step with the following characteristics :

- In each step, the volume of the crystal-like cell remains fixed. Variations of crystal-like cell parameters are not allowed during the course of minimisation.

- The six spatial parameters (coordinates and eulerian angles) for the four molecules are allowed to vary during the course of minimisation.

- The central unit cell is surrounded by two shelves in the b and c directions and one in the a direction (the interactions between 420 molecules are thus taken into consideration).

3.3 Method of calculation and results

The p-diclorobenzene has been the subject of a number of calculations by atom-atom potential method, which have been recently reviewed (ref. 10) and compared with relative stability order for polymorphs. Nevertheless, the problem of parametres for chlorine atom in these atom-atom potential methods is not trivial according to Reynolds due to the role of several lone pairs of electrons leading to anisotropic interactions, whereas atom-atom potential standard formula are purely isotropic. In this paper, the formalism of Claverie (ref. 14) has been used with the following set of parameters for chlorine atom : $k(\text{Cl})=1.1$ $R(\text{vdw})=1.76\text{\AA}$. For these calculations, we have taken X-ray data for geometric characteristics and ab-initio STO-3G charges. An activation energy of 7.5 kcal/mole (see Fig.2) is found for the $\gamma \leftrightarrow \alpha$ phase transition. This low potential barrier agrees with the ease of the phase transition.

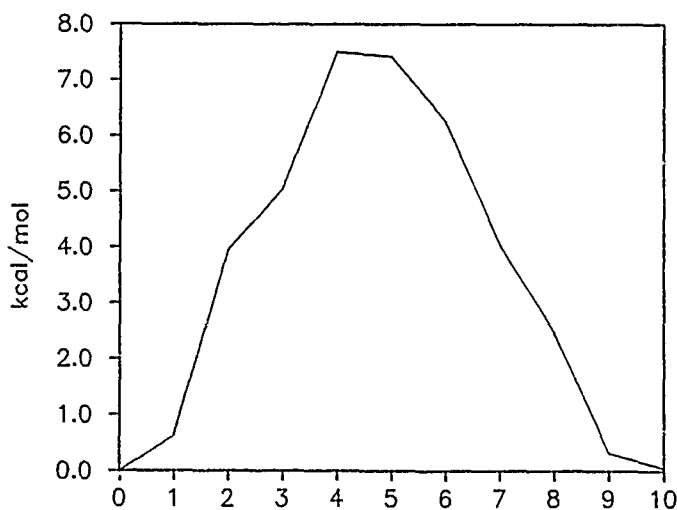


Fig. 2. Variation of packing energy along the pathway simulating the nucleus formation.

CONCLUSION

In this work we have considered the nucleus formation as cooperative displacements of molecules. The calculated low energy barrier is in agreement with the ease with which p-dichlorobenzene undergoes the phase transition between gamma and alpha forms. Our aim in the future is to gain more insight in the determination of the critical size of nucleus, determined by a balance between surface energy of the nucleus and bulk energy due to the formation of product phase.

REFERENCES

- 1 A. Burger, Pharmacy International, May (1982) 158.
- 2 L.J. Soltzberg, Susan M. Cannon, Yon Won Ho, Ellen C. Armstrong and Sophie J. Bobrowski, J. Chem. Phys. 75 (1981) 859.
- 3 D.V. Hanson, J. Chem. Phys. 63 (1975) 5046.
- 4 M.T. Dove and A.I.M. Rae, J. Chem. Soc., Faraday Disc. 69 (1980) 98.
- 5 M. Rey-Lafon and R. Lagnier, Mol. Cryst. Liq. Cryst. 32 (1976) 13.
- 6 Y. V. Mnyukh, Mol. Cryst. Liq. Cryst. 52 (1979) 163.
- 7 C.A. Fyfe and D.H. Smith, J. Chem. Soc. Faraday Trans. 2, 71 (1975) 967; A. Gavezzotti and M. Simonetta, Acta Cryst., A31 (1975) 645; P.A. Reynolds, Acta Cryst., A31 (1975) 386.
- 8 K. Mirsky and M.D. Cohen, J. Chem. Soc. Faraday 2, 12 (1976) 2155.
- 9 V. Schettino and S. Califano, J. of Mol. Str., 100 (1983) 459.
- 10 A.J. Pertsin and A.I. Kitaigorodsky, The Atom-Atom Potential Method, Springer-Verlag, Berlin, 1987.
- 11 F. Bayard, C. Decoret, J. Royer, Y.G. Smeyers and J.J. Rande, Int. J. Quant. Chem., 33 (1988) 61.
- 12 P.A. Reynolds, Acta Cryst., A33 (1977) 185.
- 13 George L. Wheeler and S.D. Colson, J. Chem. Phys, 65 (1976) 1227.
- 14 J. Caillat and P. Claverie, Acta. Cryst. A31 (1975) 448.

Monte Carlo simulations of the Acetate-Guanidinium ion pair in water

Stéphane Boudon *, Georges Wipff
Institut de Chimie ; 4, rue B.Pascal , Strasbourg 67000 France

Abstract

Monte Carlo simulations have been carried out on the guanidinium/acetate ion pair in water at 25°C and 1 atm. The free energy profile of the separation of the two ions has been calculated using the statistical perturbation theory. The determined potential of mean force between the two ions set in the configuration giving the best interactions (double-bound C_{2v} configuration) shows three minima corresponding respectively to an intimate pair, solvent separated and "infinitely" separated ion pairs. There is no clear preference for the intimate pair ($C\cdots C = 3.3\text{\AA}$). The intermediate state ($C\cdots C = 6.3\text{\AA}$) displays cooperative binding between the ions via two water molecules forming a double hydrogen-bond bridge. The two extreme states ($C\cdots C = 3.3\text{\AA}$ and 8.7\AA respectively) are separated by a large energetical barrier. The computed results provide insights in the effects of the solvent on an important biochemical system involving arginine and aspartic or glutamic acid residues and on anchoring sites of proteins.

Although electrostatic interactions in proteins are recognized as being a major driving force to determine their structure and function (1), it is still a problem to account for the "effective" attraction between charged groups in a given molecular environment. In this paper we report a theoretical study of the free energy profile (potential of mean force)(2) for the guanidinium/acetate ion pair (Gu^+AcO^-) in water to model putative interactions between solvated arginine and glutamic or aspartic acid residues. In fact because peptides involving these charged groups might act as binding or recognition sites at the surface of adhesive proteins (3), we felt important to gain insight into the solvation effects on these polar groups. Indeed, whereas the "gas phase" (Gu^+AcO^-) pair, as computed by *ab-initio* MO-LCAO-SCF calculations, is firmly bound by about -130 kcal/mol (4), solvation effects could reduce or even prevent the ion pairing.

Statistical mechanics simulations were carried out for the pair in water at 25°C and 1 atm. using the isothermal-isobaric (N,P,T) ensemble (5). The system consisted of 390 water molecules plus the ion pair in a rectangular box with periodic conditions (20x20x30Å). The Monte Carlo Metropolis (MC) procedure was used (6) modified by preferential sampling (7-8). The solute-solvent statistics were enhanced by trying to move the solute every 60 configurations. The volume changes were attempted on every 2500th configuration and involved scaling all the intermolecular separations. Configurations were generated by moving a water molecule or the solute (Gu^+AcO^-) randomly in all three Cartesian directions and by rotating it about a randomly chosen axis (9). The atom-atom interactions of the system (10-11) were calculated with a 9.0Å cutoff. The MC runs involved equilibration over 10^6 configurations and averaging over 2.4×10^6 configurations for each step

of the simulation. The free energy profile for the (Gu^+AcO^-) ion pair is computed in the context of the statistical perturbation theory (12).

In the simulation, the ion-pair is maintained in a C_{2v} symmetry and the perturbation is done along the carbon-carbon ($\text{C}\cdots\text{C}$) distance of the ions running from 3.30 to 8.70 Å by steps of 0.15 Å. Double wide sampling was used so that 0.3 Å could be covered in one MC simulation. We realized thirty-seven simulations with overlapping windows. At each step the change in free energy is evaluated as the difference in ion-water interactions and give the variation in free energy of hydration. The total energy change is computed by adding the cumulative difference in free energy to the gas-phase interaction energy between the two ions.

The calculated potentials of mean force are displayed in Figure 1. for forward and backward simulation. The fluctuations (σ) in the free energy changes averaged 0.17 kcal/mol per step. The hysteresis of the perturbation done by bringing the ions apart (forward simulation) or together (backward simulation) is small and averaged 0.84 kcal/mol. The endpoint of the pmf curve at a 8.70 Å distance is taken as reference of the perturbation energy.

Although the uncertainties in the results are difficult to establish unequivocally, they show two well-defined minima. A sharp minimum for the contact ion pair occurs at a $\text{C}\cdots\text{C}$ distance of 3.75 Å, the $\text{NH}\cdots\text{OC}$ distance is then of 1.55 Å (13). A second minimum occurs at a $\text{C}\cdots\text{C}$ separation of 6.45 Å and corresponds to a solvent separated ion pair (Figure 2.). The difference in energy between the contact ion pair and this solvent separated ion pair is weak about 1.0 kcal/mol (14). The barrier between the two states is calculated to be 11.0 kcal/mol (15). A small barrier (approximately 2.0 kcal/mol) has to be overruled to separate the ions from the solvent separated state, i.e. to disrupt the cooperative binding. A close analysis of the various configurations generated at 6.45-6.60 Å separation shows that two water molecules are

cooperatively bound between the two ions (Figure 2.) and have significant reduced mobility compared to the other water molecules in the first hydration shell.

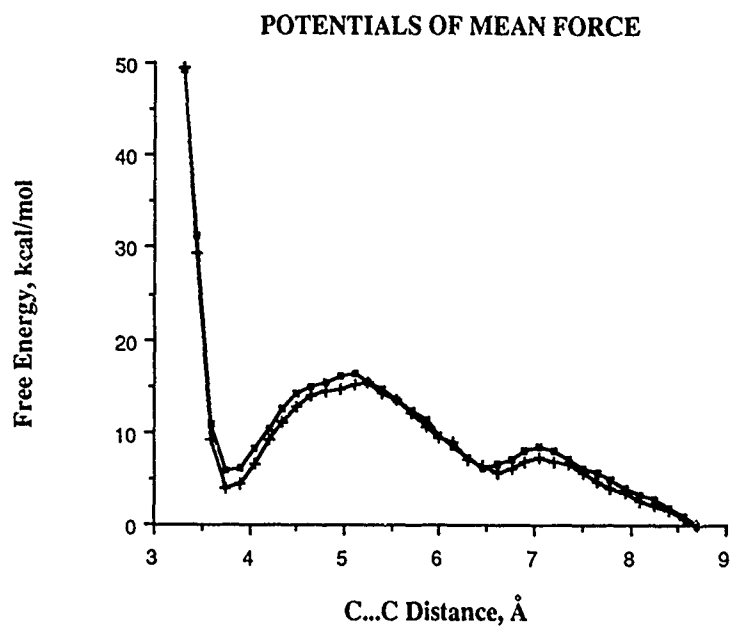


Figure 1.
Calculated Potential of Mean Force (kcal/mol.) for the acetate/guanidinium ion pair in water along the C_{2v} axis

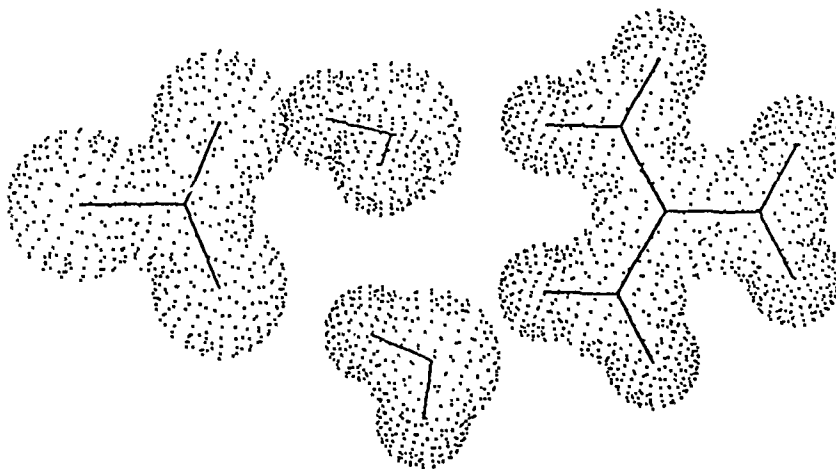


Figure 2.
A typical configuration showing the cooperative binding of two water molecules with the ions at a C...C distance of 6.45 Å

This calculated energy profile for the (Gu^+AcO^-) pair differs qualitatively from that obtained by similar computational techniques for more "spherical" ions in water (2). Indeed, the intimate or solvent separated pairs appear as local minima and are unstable towards dissociation. At a C...C distance of 8.70Å we find that the solvation free energy is still decreasing. Although some bias may have been introduced in the sampling procedure or via the cutoff distance, this may be a feature of this particular system as well.

In conclusion, we have determined the free energy profile for the guanidinium/acetate ions which shows no clear tendency for ion pairing in solution. Our results have bearing on the formation of salt bridges in proteins⁽¹⁶⁾. They suggest that polar recognition sites involving guanidinium/carboxylate pairing at the surface of adhesine proteins will be most effective if hydration is prevented (e.g. by lack of accessibility), or/and if additional stabilization results from local topology of the electric field. Although the C_{2v} arrangement is at minimum energy in the gas phase, (Gu^+AcO^-) pairing in proteins can adopt different arrangements⁽¹⁶⁾. We are investigating actually this point.

Acknowledgments

Gratitude is expressed to Dr. B. Maigret for allowance of generous computer time within the "Groupement Scientifique" IBM-CNRS and to Prof. J.M. Lehn for helpful suggestions. S. Boudon greatly acknowledges the Rhône-Poulenc Company for generous support and Prof. W.L. Jorgensen for giving the BOSS Computer Program.

References

- (1) (a) Warshel, A.; *Acc. Chem. Res.* **1981**, 14, 284 (b) Weiner, P.K.; Langridge, R.; Blaney, J.M.; Schaeffer, R.; Kollman, P.A. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, 79, 3754 (c) Van Belle, D.; Couplet, I.; Prevost, M.; Wodak, J.S.; *J. Mol. Biol.* **1987**, 198, 721
- (2) (a) Patey, G.N.; Valleau, J.P. *J. Chem. Phys.* **1975**, 63, 2334 (b) McDonald, I.R.; Rasaiah, J.C. *Chem. Phys. Lett.* **1975**, 34, 382 (c) Pangali, C.; Rao, M.; Berne, B.J. "Computer Modelling of Matter" *ACS Symp. Ser.* **1978**, 86, 32 (d) Pangali, C.; Rao, M.; Berne, B.J. *J. Chem. Phys.* **1979**, 71, 2975 (e) Hirata, F.; Rossky, P.J.; Pettitt, B.M. *J. Chem. Phys.* **1983**, 76, 4135 (f) Patey, G.N.; Carnie, S.L. *J. Chem. Phys.* **1983**, 78, 5183 (g) Kusalik, P.G.; Patey, G.N. *J. Chem. Phys.* **1983**, 79, 4468 (h) Berkowitz, M.; Karim, O.A.; Mc Cammon, J.A.; Rossky, P.J. *Chem. Phys. Lett.* **1984**, 105, 577 (i) Ravishanker, G.; Beveridge, D.L. *J. Amer. Chem. Soc.* **1985**, 107, 2565 (j) Pettitt, B.M.; Rossky, P.J. *J. Chem. Phys.* **1986**, 84, 5836 (k) Dang, L.X.; Pettitt, B.M. *J. Chem. Phys.* **1987**, 86, 6560 (l) Jorgensen, W.L.; Buckner, J.K.; Huston, S.E.; Rossky, P.J. *J. Amer. Chem. Soc.* **1987**, 109, 1891 (m) Dang, L.X.; Pettitt, B.M. *J. Amer. Chem. Soc.* **1987**, 109, 5531 (n) Buckner, K.; Jorgensen, W.L. *J. Amer. Chem. Soc.* **1989**, 111, 2507 (o) Jorgensen, W.L. *J. Amer. Chem. Soc.* **1989**, 111, 3770 (p) Jorgensen, W.L. *Acc. Chem. Res.* **1989**, 111, 3770
- (3) (a) Piersbacher, M.D.; Hayman, E.G.; Ruoslahti, E. *Proc. Natl. Acad. Sci. USA* **1983**, 80, 1224 (b) Piersbacher, M.D.; Ruoslahti, E. *Proc. Natl. Acad. Sci. USA* **1984**, 81, 5985 (c) Ruoslahti, E.; Piersbacher, E. *Cell* **1986**, 44, 517 (d) Naidet, C.; Sémériva, M.; Yamada, K.M.; Thiery, J.P.; *Nature* **1987**, 325, 348 (e) D'Souza, S.E.; Ginsberg, M.H.; Burke, T.A.; Lam, S.C.-T.; Plow, E.F.; *Science* **1988**, 242, 91
- (4) (a) Sapse, A.M.; Russel, C.S. *Int. J. Quant. Chem* **1984**, 26, 91 (b) Sapse, A.M.; Russel, C.S.; *J. Mol. Struct.* **1986**, 137, 43
- (5) McDonald, I.R. *Mol. Phys.* **1972**, 23, 41
- (6) Metropolis, N.; Rosenbluth, A.W.; Rosenbluth, M.N.; Teller, A.H. *J. Chem. Phys.* **1953**, 21, 1087
- (7) (a) Owicki, J.C.; Scheraga, H.A.; *Chem. Phys. Lett.* **1979**, 47, 600 (b) Bigot, B.; Jorgensen, W.L.; *J. Chem. Phys.* **1981**, 74, 1944
- (8) A $1/(r^2+c)$ weighting factor was chosen where $c=500\text{\AA}^2$ in order to have an adequate sampling
- (9) The ranges for these motions were chosen to provide acceptance probabilities of ca. 40% for new configurations and were 0.15\AA and 15° for water monomers, 0.06\AA and 6° for the solutes.
- (10) The geometry of the ions has been determined from *ab-initio* MO-LCAO-SCF calculations using the extended 6-31G(d) basis set. The intermolecular potential function is pairwise additive-including Coulomb and Lennard-Jones terms-and do neither take into account higher order effects due to three body interactions nor polarization effects. Parameters for acetate and guanidinium, were developed through *ab-initio* MO-LCAO-SCF calculations with the 6-31G(d) basis set for low energy forms of complexes between water and ions. The TIP4P model of water was used in conjunction with the above described parameters.
- (11) Jorgensen, W.L.; Tirado-Rives, J. *J. Amer. Chem. Soc.* **1988**, 110, 1657

(12) (a) Zwanzig, R.W. *J. Chem. Phys.* **1954**, 22, 1420 (b) Jorgensen, W.L.; Buckner, J.K.; Huston, S.E.; Rossy, P.J. *J. Amer. Chem. Soc.* **1987**, 109, 1891

(13) In agreement with the *ab-initio* calculations of reference 2a within 0.1 Å

(14) 0.4 kcal/mol when the perturbation is forward and 1.6 kcal/mol when the perturbation is backward

(15) 10.6 kcal/mol going forward, and 11.4 kcal/mol going backward

(16) Singh, J.; Thornton, J.M.; Snarey, M.; Campbell, S.F. *FEBS Lett.* **1987**, 224, 161

SIMULATION OF POLYIONIC INTERACTIONS BY PROGRESSIVE APPROXIMATIONS

Heiko H. OHLENBUSCH

Institut de Physique Biologique, 4 rue Kirschleger,
F-67085 Strasbourg (France)

SUMMARY

Computer generated ionic site correlation functions were applied to polyionic macromolecules in order to obtain information about the relative probability of stereospecific interactions between them. The results show that it is possible to predict preferred mutual orientations of polyions and to estimate the most probable path of approach of interacting charged particles. The method of progressive approximations can be applied step-wise to long-range electrostatic interactions. Application of the analysis to highly charged model systems shows that it is possible to select geometric correlations between alternative polyionic conformations by neglecting most of the short-range molecular forces.

INTRODUCTION

The method of progressive approximations is implicit in most theoretical approaches to gain insight into molecular interactions and it has been used before for the estimation of interactions between biopolymers (ref. 1). The problems encountered in analyzing interactions between macromolecules are so formidable, that approximations are essential. Prior attempts to explain and predict polyion interactions involved extensions of the Debye-Hueckel theory for ionic solutions to various macromolecular model systems (ref. 1) with the primary aim being the derivation of their thermodynamic parameters in idealized solutions. Various types of symmetry were assumed to permit the use of geometric operators, in order to obtain expressions in closed form, which could be solved by the methods of calculation available at the time.

Macromolecules of biological interest, however, are surrounded by such a variety of components and confined to numerous types of cellular compartments, the structures of which are still not sufficiently well understood, that the question of the relevance of extrapolating thermodynamic data obtained from solutions to in vivo conditions in order to explain their biological behavior is difficult to answer. Of more immediate success may be the investigation of structural correlations using distance geometry independent of detailed thermodynamic considerations, to avoid complications at least in the initial stages of the calculations. The advent of fast computers made possible such an alternative approach to determine the probability of macromolecular complex formation (ref. 2) and in recent years the crucial importance of electrostatic interactions for the recognition between charged macromolecules is receiving renewed attention (ref. 3).

The use and analysis of simple, fly-by ionic site correlation curves, obtained with polyionic model structures permit insight into the relative probability of structural complementarity, which is of great importance for understanding the role of charged biological macromolecules like proteins, nucleic acids, phospholipids etc.

METHOD

The approach using distance geometry starts by assuming that the distribution of point charges located on macro-ions is known and asks about the relative probability of fitting these charge assemblies to each other or to other known structures. The fitting probabilities are then used for the selection of reference structures, which in turn serve for the generation of new sets of model configurations of polyions by relaxing more and more of the initial approximations made to represent the various ionic sites.

In general, any type of site correlation function used to estimate the strength of molecular interactions varies inversely with distance and depends quite naturally on the type of interacting sites investigated: van der Waals interactions are estimated from functions of the type $f(1/R_{ij}^6)$, interactions in the presence of counterions from $f(1/(R_{ij} \exp(k \cdot R_{ij})))$, the electrostatic interactions from the strength of the electrostatic fields, i.e. $f(1/(R_{ij} \cdot R_{ij}))$ etc., where R_{ij} are the interatomic distances between atoms i and j and k the Debye-Hueckel parameter (ref. 1).

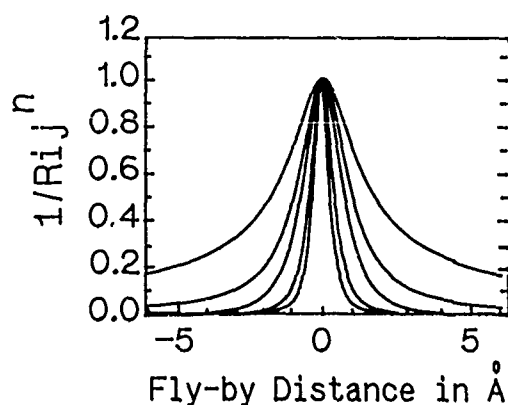


Fig. 1. Plots of $1/R_{ij}^n$ for a positive charge moving along the X-axis at a relative distance Y , set equal to 1, past a negative charge fixed at $X = 0$, $Y = 0$ are superimposed to show that the precision of site location does not depend very much on the value of the exponent $n = 1, 2, 3, 6$ or 12 for the 5 curves.

Fig. 1 summarizes a few selected fly-by correlation curves in order to show that the choice of the exact type of site correlation function employed for the estimation of structural parameters is not of primary importance (see also ref.

2). The localization of a charged site by a moving probe does not depend very much on the actual value of the exponent of R_{ij} , as long as it is negative and its absolute value is not much smaller than 1. The method of fly-by ionic site correlation analysis was developed in order to investigate the probability of long-range interactions and to select preferred mutual orientations of approaching charged macromolecules (ref.2). For the present analysis the exponent of R_{ij} was set equal to -2. To account for charge attraction or repulsion, individual R_{ij} values were multiplied by S_{ij} equal to +1 or -1, respectively.

As a first approximation the analysis is applied to a number of standard geometric model structures and asks whether these structures have a propensity to fit either to each other or to other structures known to be present in their vicinity. For instance, any known protein sequence can be analyzed for the presence of "one-dimensional" structural complementarity by assuming that each amino acid has associated with it one characteristic property and then investigating whether complementary patterns can be detected in the sequence. Thus, the probability, that a polyion like poly-arginine interacts intermolecularly with itself, is much lower than that of a mixed polymer of poly-arginine and polyglutamic acid. In going from the 1D approximation to 2D or 3D, one can determine, whether charged residues, mutually attractive in 1D, are distributed in a pattern such that their charges are oriented along a particular radial direction when folded into an alpha-helix (Fig. 2) or along one particular side of a 2D

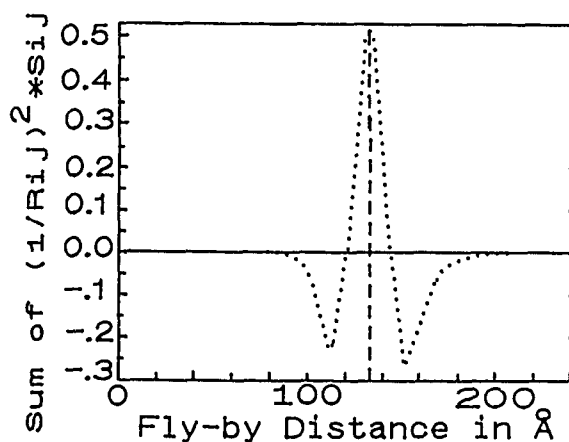


Fig. 2. Plot of $(1/R_{ij}^2) * S_{ij}$ for a hypothetical, alpha-helical probe of poly-Ala with an internal section of R(19-27)E(28-36) moving past an identical molecule which is oriented antiparallel to the probe in a coordinate system. The fly-by correlation curve shows one maximum of attraction and 2 minima of repulsion and permits the pairwise alignment of the identical molecules corresponding to the interaction maximum. This alignment determines the position of the dyad axis of symmetry of the polyionic complex.

beta-sheet model structure or not. Any increase in the probability of polyionic neutralization would suggest, that the structures involved would be stabilized by such an interaction and thus selected out from the large variety of available competing configurations. Examples of some of these analyses are shown in Fig. 2 to 4.

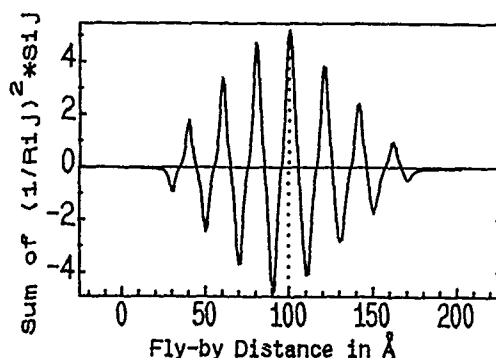


Fig. 3. Fly-by correlation curve of an artificial polyion, polyalanine (Ala)50 with blocked ends and Glu at positions 19,26,33,40 and Arg at positions 22,29,36, and 43, designed to mimic the central segment of histone H4 as alpha-helix. The dotted line, passing through the maximum of the correlation curve, indicates the position of maximum intermolecular charge neutralization and locates the dyad axis of the corresponding intermolecular dimer with high precision. Note that the distances separating all peaks are close to 20 Å, i.e. a distance equal to the diameter of a normal B-DNA double-helix.

To summarize, a general approach to investigating polyionic interactions of linear charged macromolecules would involve the following steps:

- 1.0. Representation of the primary monomer sequence by a one-dimensional string of equi-distant points.
- 1.1. The 1D analysis may be improved, by imposing that the distances separating the points correspond to the known atomic backbone separations between monomer groups.
- 1.2. A further improvement in the model would be the replacement of the points by polygons or spheres with residue-specific radii, so that their volumes equal those of the van der Waals volumes of the particular monomers.
- 2.0. The next step would extend the compressed information of the string model into two directions, for instance by assuming for a peptide, that its primary sequence is folded into a beta-sheet. In this case the properties of the primary sequence could be represented by points located at the positions of the corresponding alpha-carbons as the initial 2D structural approximation.
- 2.1. The 2D analysis could be improved, by imposing that the points at the alpha-carbon positions become polygons with dimensions determined as above from the van der Waals volumes of the corresponding residues.
- 3.0. The next step would concern the representation of primary sequences in the

form of idealized 3D configurations of known dimensions. Thus, DNA sequences may be folded into paired or unpaired helices of particular conformations, in order to simulate native or denatured DNA or such structures as B-DNA, A-DNA, C-DNA etc. Similarly, proteins may be folded into ideal alpha-helices or beta-sheets or beta-turns etc. In general, the 3D analysis should start with those structures, which are known to be the dominating structures for the case of interest. For instance, there is little incentive to study the behavior of beta-sheets for histone proteins, which are known to exist to 55 \pm 5 % as alpha-helices in native chromatin and do not show any evidence of beta-sheet regions in the presence of DNA (ref. 4-5).

Thus, 3D approximations for proteins rich in alpha-helices would start by representing the amino acid sequence as cylinders of dimensions corresponding to that of an alpha-helix, i.e. a radius of 2.3 A and a length equal to $(N-1) \times 1.47$ A, where N is the number of residues in the sequence. The residue properties would be concentrated onto points on the cylinder surface corresponding to the locations of the alpha-carbon atoms.

- 3.1. Similar to the approximations made for the 1D or 2D models, the next step would involve the expansion of the points to polygons with residue-specific dimensions and properties spread over the external surface.
- 3.2. The next step would replace the polygons by rigid, radially extended residues with their properties concentrated on points located at corresponding atomic positions.
- 3.3. The next step would replace the points by the actual atoms of the residues surrounded by van der Waals spheres.
- 3.4. The rigidity of the residues could be relaxed in the next approximation step, either by replacing them by time-averaged rotamers or other models aimed at representing their dynamic properties.

Similarly, 3D approximations for DNA sequences would start by representing the base pairs as cylinders of dimensions corresponding to those of the B-DNA double-helix, i.e. a radius of 10.2 A and a length equal to $(N-1) \times 3.38$ A, where N is the number of basepairs. The base pair properties would be concentrated onto points on the cylinder surface corresponding, for instance, to the locations of the phosphate atoms. Steps 3.1 to 3.4 would be similar to those for proteins, except that the base pairs could be better simulated by cylinder slices rather than linear, radially extended strings.

Many further steps of successive approximations may be required before a satisfactory answer is found for the prediction of biological structures in the in vivo environment, but it was very instructive to find, that already the initial steps of the analysis using very simple approximations as outlined above gave conclusive geometric parameters for molecular alignment. This observation is important, since the higher the relevant information content of simple approximations, the lower is the effort needed to understand the phenomena under consideration.

For instance, the possibility to estimate long-range interactions between protein alpha-helices correctly by considering them as idealized, rigid cyclinders, would make it considerably easier to investigate the behavior of the remaining peptide sequences. The possibility of selecting those peptides, which bind stereo-specifically into one or the other of the two principal grooves of double-helical DNA as suggested (ref. 9-11), would considerably facilitate the search for reasonable configurations for any adjacent, less rigidly bound peptide regions.

Thus, one of the essential features of the procedure of successive, step-by-step approximations of polyion interactions concerns the analysis of the relevance of the results to the problem at hand after each of the analysis steps. The data presented in Fig. 4 may be used to illustrate the initial step of such an analysis. Supposing that it is possible to rationalize the effect of diamines on the protection of DNA against melting by the very rudimentary model used, i. e. consisting of only 2 negatively charged phosphate sites and 2 positively charged amino groups located in a vacuum, then this observation, derived from the first step of the approximation, implies that simple coulombic interactions are the main driving force of the observed phenomenon. The analysis predicts correctly, that the diamine with 5 C-atoms (i. e. cadaverine) interacts most strongly with DNA and protects it most efficiently against heat denaturation. It shows furthermore, that this prediction is due to the correspondence of the distance between phosphates along the backbone of the B-form DNA, which is 6.854 Å, and the distance between the 2 N-atoms of extended cadaverine molecules. Furthermore, the simulation shows that selectivity between diamines is large enough to differentiate between diamines, which differ by only one C-atom, which is equivalent to a difference in the N-N distance of only about 0.88 Å, if it is assumed that the diamines are in their extended configuration. That this is a very plausible assumption is confirmed by the drop in DNA protection by diamines with more than 5 C-atoms, which shows that the amino groups remain maximally separated and that the carbon-skeleton does not bend significantly to generate a conformation which would bring the two amino groups close enough to each other, so that their separation matches that of the phosphates of the DNA backbone.

That the data can be simulated at such a low level of sophistication, i.e. without taking into account solvent and counterion effects, partial charges or charge fluctuations etc., may be due to fortuitous averaging or compensating effects or it may indicate that these parameters are of minor importance. Only additional model experiments specifically aimed at resolving such questions will show, which of the possible answers are relevant.

A more detailed investigation could be aimed at simulating the effects of DNA base composition and/or base sequence on the melting data in the presence of diamines or at trying to understand in detail, why solvent and partial charges do not need to be taken into account in order to derive basic structural correlations about overall molecular complementarity between charged molecules. The effect of varying some of the critical parameters, like distance of closest

approach between molecules, the value of the exponent used for the distance correlation, the effect of modifying the values of +1 and -1 assigned to the point charges etc., used for the calculation of the fly-by correlation curves may also be instructive and may highlight some of the additional applications of the method.

The answer concerning the relevance of the results is important for the selection of the next step towards rendering the approximations more realistic. With the accumulation of experience in the successful prediction of experimental data, a set of criteria will evolve, which can in turn be used to improve the success of predictions in ever greater detail.

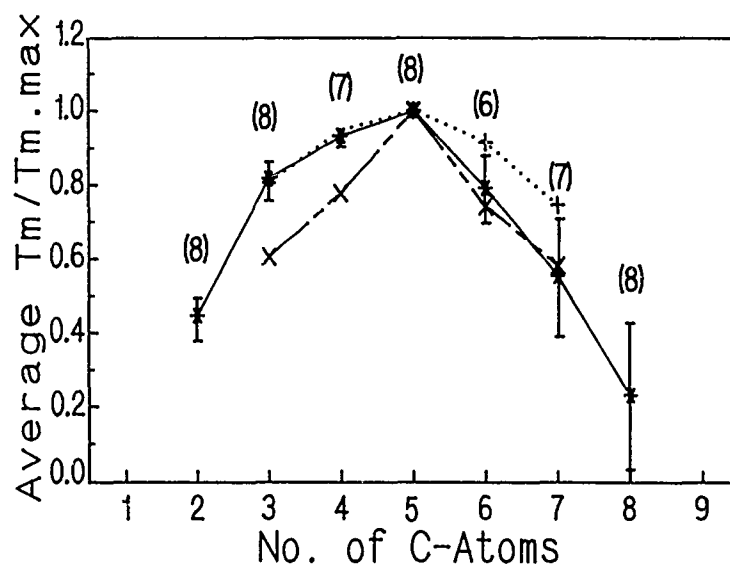


Fig. 4. Superposition of experimental and theoretical data, to demonstrate typical applications of the methodology described. Curve 4A (—) represents the relative change in the midpoint melting temperature, T_m , of DNA in the presence of linear diamines of different C-content. The Y-values were obtained by averaging the T_m data over the number (in parentheses) of the DNA samples used (ref. 7) and by normalizing them to 1 for the maximum average T_m observed. The standard error bars indicate primarily the dependence of T_m on the differing GC-content of the DNA samples used. The X-values represent the number of diamine C-atoms separating the terminal, positively charged amino groups. Curves 4B (...) and 4C (---) represents the relative change of the maximum Y-values obtained from fly-by correlation curves of various diamines interacting with a fixed diphosphate with a charge separation equal to that of the backbone of B-DNA, i.e. 6.854 Å. The diamines were simulated by positive point charges separated by distances corresponding to those of the extended C-backbone. Curve 4B results from a fly-by at Y = 1 Å and 4C at Y = 2 Å.

CONCLUSION:

The results presented show that the actual type of correlation function used to analyze long-range, inter-molecular ionic interactions is of only minor importance, because of their relative insensitivity to the value of the exponent of the distance parameter (Fig. 1). This in turn suggests, that the effects due to solvent interactions, counterions and electric polarization are dominated by long-range electrostatic interactions as far as overall conformational selection is concerned, when the molecules, taking part in the interactions, are highly charged species. That this is indeed the case becomes clear, for instance, by comparing the melting data of DNA in the presence of various diamines (ref. 7) with data obtained from ionic site correlation curves (Fig. 4).

The results show that fly-by ionic site correlation curves are very sensitive to the location of charged sites as well as to the amount and sign of the charges present. This observation makes possible the estimation of interaction maxima and minima as a function of molecular configuration and orientation and thus of the preferred mutual orientations of interacting polyionic species. The method of analysis has been applied with encouraging results to complex, highly charged macromolecules, like for instance histones, the basic nuclear proteins, as well as acidic polynucleotides like DNA, in order to test the self-consistency of the high-resolution molecular model for chromatin subunits (ref. 6), which is now available as a colored, three-dimensional display by computer graphics (ref. 8).

REFERENCES

- 1 C. Tanford, Physical Chemistry of Macromolecules, John Wiley, New York, 1961, pp. 457-525.
- 2 H. H. OHLENBUSCH, Polyionic Interactions and Structural Complementarity in Model Histone-DNA Complexes, Symposium sur les Polyélectrolytes et Interfaces Chargées, Strasbourg 10.-11. July 1981.
- 3 N.G.J. Richards, Electrostatics and Molecular Recognition, Int. J. Quant. Chem. Suppl. 15 (1988) 85-93.
- 4 G.J.Jr. Thomas, B. Prescott and D.E. Olins, Science 197 (1977) 385-387.
- 5 N.A. Nicola, A.W. Fulmer, A.M. Schwartz and G.D. Fasman, Biochemistry 17 (1978) 1779-1785.
- 6 H.H. Ohlenbusch, Structural Complementarity in Chromatin Subunits, Naturwissenschaften 66 (1979) 206-207.
- 7 H.R. Mahler and B.D. Mehrotra, The interaction of nucleic acids with diamines, Biochim. Biophys. Acta 68 (1963) 211-233.
- 8 H.H. Ohlenbusch, Computer Graphics, European Molecular Biology Laboratory, Postfach 10.2209, D-6900 Heidelberg 1, Germany.
- 9 T.Y. Shih and J. Bonner, J. Mol. Biol. 48 (1970) 469-487.
- 10 M.T. Sung and G.H. Dixon, Proc. Nat. Acad. Sci. 67 (1970) 1616-1623.
- 11 H.H. Ohlenbusch, Structural complementarity of histone 4 antiparallel dimers with DNA, Spec. Sci. Tech. 4 (1981) 359-366.

**HIGH TEMPERATURE ANNEALED MOLECULAR DYNAMICS
SIMULATIONS AS A TOOL FOR CONFORMATIONAL
SAMPLING:
A TEST CASE ON THE 222 BICYCLIC CRYPTAND. †**

P. AUFFINGER, G. WIPFF*

Laboratoire de RMN et de Modélisation Moléculaire
Institut de Chimie
4 rue Blaise-Pascal
67000 STRASBOURG

ABSTRACT:

We have performed methodological studies using "High Temperature Annealed Molecular Dynamics Simulations" (HTAMDS) on a bicyclic cryptand, with the following protocol : 100 ps of molecular dynamics at 1000°K, followed by optimisation of the structures saved every 0.2 ps, relaxation of these structures during 20 ps of molecular dynamics at 300°K and, finally, optimisation. Four sets of 500 low energy conformers of the free 222 cryptand have been produced starting either from the free cryptand or from the M⁺/222 cryptate with different representation of M⁺.

The analysis of these four sets allows assessment of the ability of this HTAMDS technique : (i) to interconvert experimentally known conformers starting from one of them, (ii) to locate the energy minima, (iii) to generate new conformers of low energy, and (iv) to account for the average structure observed on the NMR time scale. In view of the ionophoric behavior of the cryptand, structures are analyzed in terms of the "in/out" orientation of the binding sites.

It is found that the annealed simulations on the free molecule, although sampling largely the conformational space, do not give structures adequate for cation inclusion. They generate the lowest energy structure known experimentally and other new closely related ones. Inclusion of cation in the simulation (either as a purely electrostatic "driver", or as a charged sphere) leads to conformations found in several complexes.

In terms of "receptor design" it is thus essential to consider explicitly the complex formed with the host in order to find conformations suitable for binding. Conversely, in the field of "drug design", conformations of the drug recognized by its receptor may not be found by HTAMDS if the receptor (generally of unknown structure) is not taken into account.

† To be published in *J. Comput. Chem.*, 1990, 11, 000.

STRUCTURAL AND CHARGE EFFECTS ON THE PHTHALOCYANINE DIMER

F. TORRENS, E. ORTÍ, and J. SANCHEZ-MARIN

Departament de Química Física, Facultat de Química, Universitat de València
 Dr. Moliner 50, 46100 Burjassot, Valencia, (Spain)

SUMMARY

The structure of the phthalocyanine dimer is determined by theoretical calculations utilizing an atomic pair potential function (R-1-4-6-12 type) proposed by S. Fraga. A face-to-face, slipped structure is reported for the dimer. Structural and oxidation state effects on the association energy are discussed. Two clear advantages appear when oxidation state effects are used for describing the phthalocyanine dimer: 1) The association energy is increased. 2) The interaction energy is less sensitive to the effect of the geometrical parameters (distance between the molecular planes and rotation angle).

INTRODUCTION

Aggregation is a well-known phenomenon in phthalocyanine chemistry. Interactions can occur between adjacent phthalocyanine rings both in organic and aqueous phase, resulting in coupling between the electronic states of two, or more, phthalocyanine units (ref. 1).

Phthalocyanine compounds, metallophthalocyanines, are known to present semiconducting properties. When metallophthalocyanines or metal-free phthalocyanine are partially oxidized with a halogen (typically I_2), they become conducting while simultaneously adopting a face-to-face stacking (ref. 2).

METHOD

A theoretical study on the intermolecular interaction between two phthalocyanine molecules has been carried out by using an atom-to-atom pair potential formulation proposed by S. Fraga (ref. 3-4). The potential is defined as:

$$\begin{aligned}
 E_{ij} = & 1389.4168 \, q_i q_j / R_{ij} \\
 & - 694.70838 \, (f_i \alpha_i q_j + f_j \alpha_j q_i) / R_{ij}^4 \\
 & - 1516.0732 \, f_i \alpha_i f_j \alpha_j / \{ (f_i \alpha_i / n_i)^{1/2} + (f_j \alpha_j / n_j)^{1/2} \} R_{ij}^6 \\
 & + 4.184 \, c_i c_j / R_{ij}^{12}
 \end{aligned} \tag{1}$$

where q (atomic net charge), f (fitting optimized factor), α (atomic

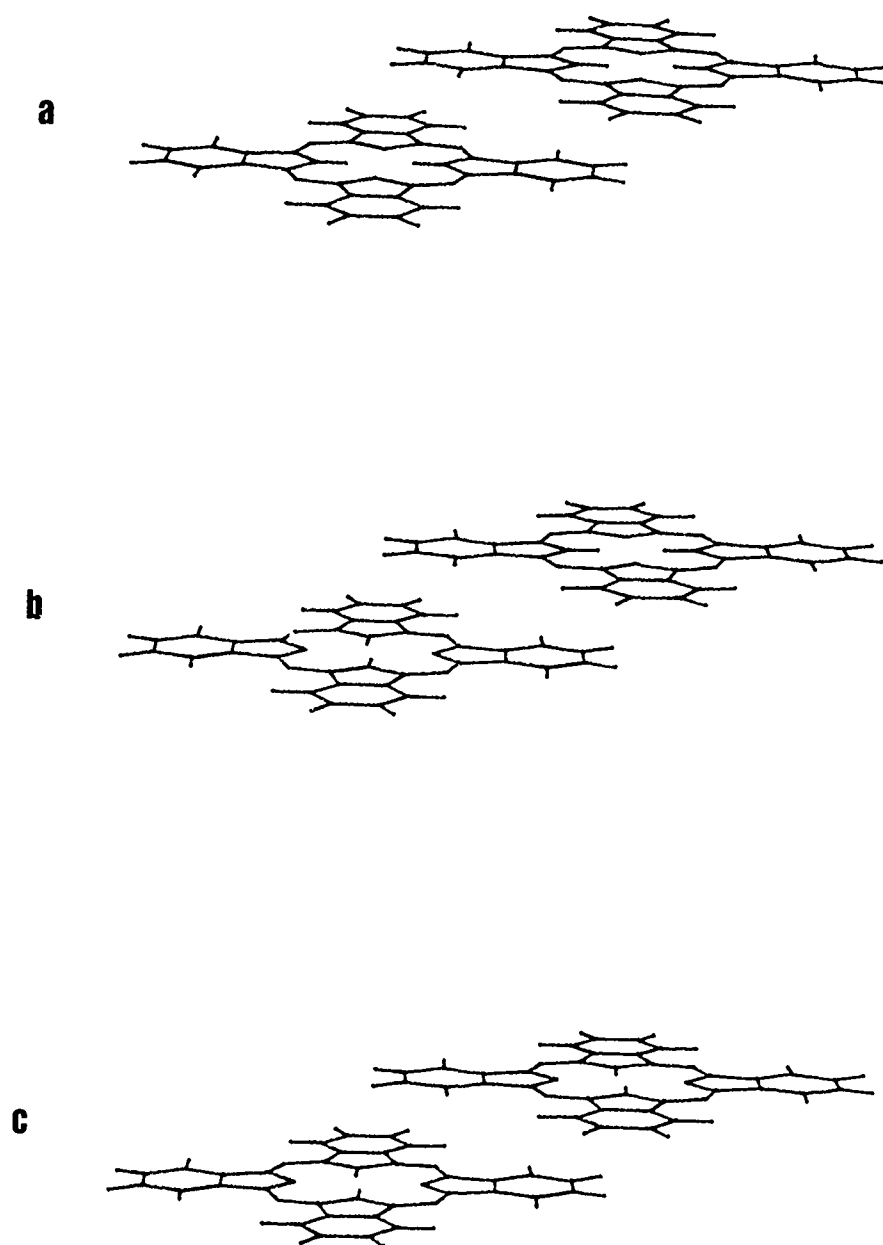


Fig. 1. Stacking face-to-face, slipped structures of the phthalocyanine dimer. A conical projection on the ZY plane is shown. The Z and Y axes denote the vertical and horizontal directions, respectively. The X axis lies on the viewing direction. Figure a (b, c) corresponds to the a (b, c) minimum in Table 1.

polarizability), n (atomic effective number of electrons), and c (overlap coefficient) are assigned to each (a,b) class of (i,j) atoms belonging to the (A,B) molecule and according with a classification given by Clementi (ref. 5).

DEVELOPMENT OF THE POTENTIAL AND NEW IMPLEMENTATIONS

An empirical correction of the dispersion energy term is included by using a new R^{-6} term (ref. 6) that is added to the original potential like a second R^{-6} term (ref. 7-9).

The R^{-4} and the two R^{-6} terms are damped by using a formula by Douketis et al. (ref. 10). This formulation is given:

$$E^{\text{DAMP}}(R) = E(R) f(R) g(R,n) \quad (2)$$

The g function is given by:

$$g(R,n) = \{1 - \exp(-2.1 R/n - 0.109 R^2/n^{1/2})\}^n \quad (3)$$

and corrects for the overlap effects. The values of n in eq. (2) are 4 and 6 for the R^{-4} and R^{-6} terms, respectively. The f function is given:

$$f(R) = 1 - R^{1.68} \exp(-0.78 R) \quad (4)$$

and corrects for the exchange effects.

Two different procedures of renormalization of the molecular electrostatic charge have been implemented: 1) Local renormalization of the electrostatic charge in a molecular fragment. 2) Global renormalization of the electrostatic charge through the overall molecule.

A variable metric algorithm has been implemented in order to optimize various starting geometries.

RESULTS

Stacking face-to-face, slipped (see Figure 1) and face-to-face (see Figure 2) configurations of the phthalocyanine dimer have been revealed as the most important orientations in the crystal structures of β -phthalocyanine (ref. 11) and halogen-doped salts (ref. 12). The association energies and distances of these minima are reported in Table 1.

In order to test the effect of the oxidation state of the phthalocyanine molecules, the structures of the dimer are optimized with various oxidation numbers. The results are written in Table 2.

A different parameterization of Fraga's potential has been also tested. "Ab initio" STO-3G net charges (ref. 13) have been used in order to calculate the R^{-1} and R^{-4} energy components in eq. (1). Atomic polarizabilities have been also interpolated for the new charges in order to calculate the R^{-4} and R^{-6} energy terms in eq. (1). The results are shown in Table 3 for neutral molecules

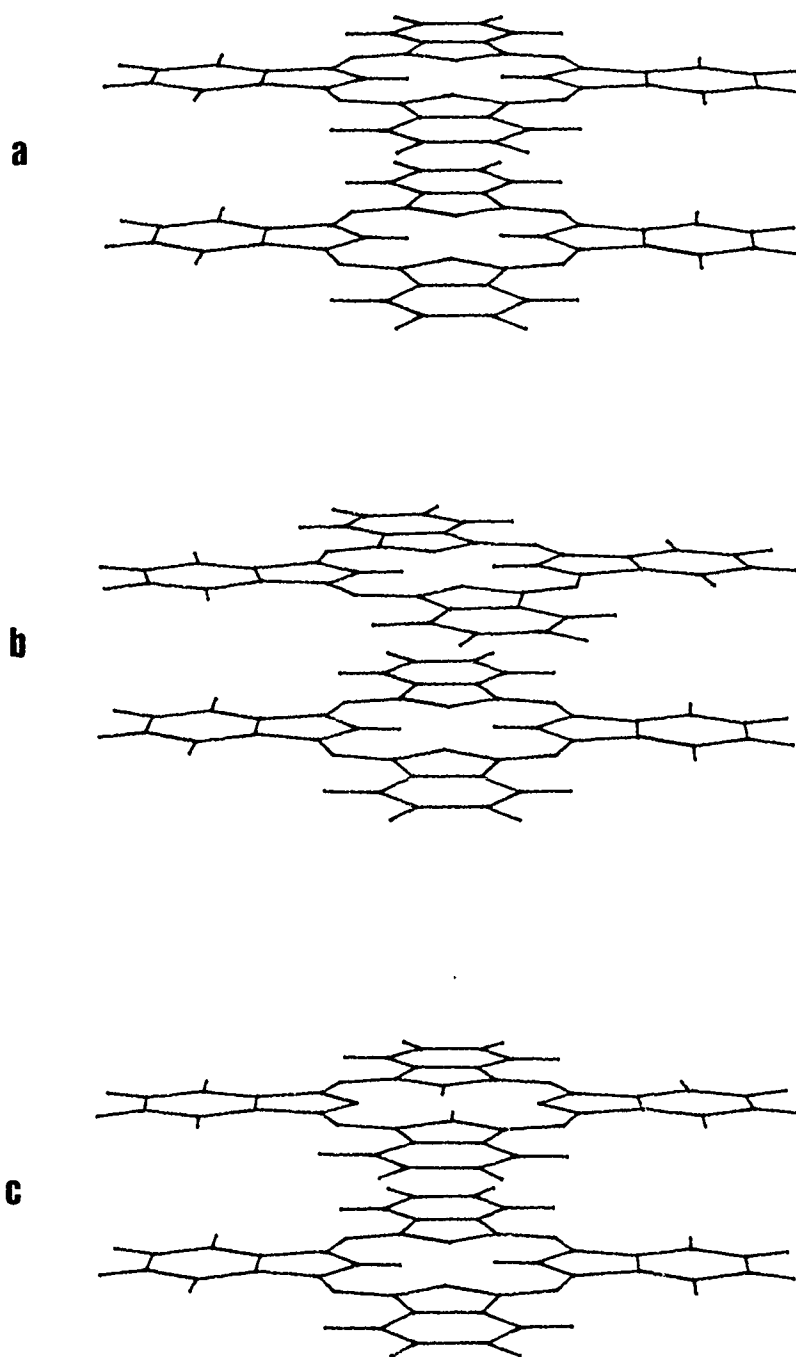


Fig. 2. Face-to-face structures of the phthalocyanine dimer. A conical projection on the ZY plane is shown. Figure a (b, c) corresponds to the d (e, f) minimum in Table 1.

and in Table 4 for various oxidation states.

TABLE 1

Energies and geometrical parameters of the phthalocyanine dimer.

minimum	type ^a	-E (kJ/mole)	horizontal slipping (Å)	vertical stacking (Å)	rotation angle (°)
a	FFS	236.0	7.8	3.2	0.0
b	FFS	233.3	7.9	3.2	93.0
c	FFS	230.6	8.0	3.1	0.0
d	FF	5.1	0.0	3.4	0.0
e	FF	6.7	0.0	3.4	10.0
f	FF	3.9	0.0	3.4	90.0

^aSymbols employed: FFS: Face-to-Face, Slipped; FF: Face-to-Face.

TABLE 2

Association energies (-E, kJ/mole) for various oxidation numbers of the phthalocyanine dimer.

minimum	type ^a	oxidation states of molecules A;B		
		0;0	1/3;1/3	0;1
a	FFS	236.0	243.6	276.1
b	FFS	233.3	240.2	273.6
c	FFS	230.6	236.9	268.4
d	FF	5.1	55.1	-*
e	FF	6.7	57.3	117.5
f	FF	3.9	54.1	116.2

^aSee footnote of Table 1.

*The trial geometry of d minimum optimizes into the final geometry of the e minimum.

DISCUSSION

Two basic stacking structures (face-to-face, slipped a-c minima and face-to-face d-f minima) are important in order to describe the phthalocyanine dimer (see Table 1). The stabilities of the slipped a-c minima are greater with difference. The horizontal slipping effects reduce the vertical distance between the molecular planes in 0.2 Å. Changes in the rotational angle yield to a small effect on the association energies.

However, the effect of the oxidation state is very significant and, when the molecular charges of the molecules are increased, the association energies are highly intensified (see Table 2). In turn, the less stable face-to-face structures are, specially, highly stabilized and the difference with the most stable minimum highly decreases.

TABLE 3

Energies and geometrical parameters of the phthalocyanine dimer. STO-3G net charges are used in eq. (1).

minimum	type ^a	-E (kJ/mole)	horizontal slipping (Å)	vertical stacking (Å)	rotation angle (°)
a	FFS	-*	-	-	-
b	FFS	-†	-	-	-
c	FFS	-*	-	-	-
d	FF	381.2	0.0	3.3	0.0
e	FF	-*	-	-	-
f	FF	388.3	0.0	3.2	90.0

^aSee footnote of Table 1.

*The trial geometries of a, c, and e minima optimize into the final geometry of the d minimum.

†The trial geometry of the b minimum optimizes into the final geometry of the f minimum.

TABLE 4

Association energies (-E, kJ/mole) for various oxidation numbers of the phthalocyanine dimer. STO-3G net charges are used in eq. (1).

minimum	type ^a	oxidation states of molecules A;B		
		0;0	1/3;1/3	0;1
a	FFS	-*	-*	-*
b	FFS	-†	-†	-†
c	FFS	-*	-*	-*
d	FF	381.2	365.9	399.8
e	FF	-*	-*	-*
f	FF	388.3	373.3	407.4

^aSee footnote of Table 1.

*The trial geometries of a, c, and e minima optimize into the final geometry of the d minimum.

†The trial geometry of the b minimum optimizes into the final geometry of the f minimum.

No improvement has been obtained with the use of "ab initio" STO-3G net charges in eq. (1). No face-to-face, slipped structure is reported in Table 3 with the employment of this parameterization. Only two stacked face-to-face (d and f) minima are described in Table 3. The effect of the oxidation state is smoothed when STO-3G net charges are used (see Table 4).

A multipolar analysis of the electrostatic properties of the phthalocyanine molecule is interesting at this point. STO-3G net charges provide a too low quadrupole moment (57.6 atomic units) when compared with the standard one

(358.4 a. u.). A low quadrupole moment favours face-to-face structures while a high quadrupole moment favours face-to-face, slipped minima, in better agreement with the crystal structure of the β -phthalocyanine (ref. 11). This effect has been previously described for the benzene, s-tetrazine, and their mixed dimer (ref. 14).

Fraga's atom-atom pair potential implemented into the AMYR program (ref. 3-4) allows for the minimization of the association energy considering the six intermolecular degrees of freedom (three translational and three rotational) of molecule B as a whole. Three derivative methods are implemented to deal with the geometrical optimization: gradient steepest-descent, rank I variable metric Davidson (ref. 15), and rank II variable metric Broyden-Fletcher-Goldfarb-Shanno (BFGS) (ref. 16) minimization methods.

Thus, Fraga's potential is a good method for evaluating analytical energies and gradients given the small number of degrees of freedom involved. It should be noted that a double addition of the terms in eq. (1) is needed, and the size of this addition is formally proportional to the square of the number of atoms in the molecules; so, a great interest exists for incorporating vectorization into this pair-potential algorithm (ref. 17).

The basic approach to exploiting vector processors is to calculate Fraga's potential interactions in a double loop over atoms of molecules A and B, storing in square matrices information about the potential such as the inverse powers of the distance and the specific energy terms in eq. (1). This is possible because of the independence of the atom-atom interaction calculations. Also, the analytical gradient of Fraga's potential has the same properties and atom-atom interaction independence.

CONCLUSION

Two clear advantages appear when oxidation state effects are used for describing the phthalocyanine dimer: 1) The association energy is increased. 2) The interaction energy is less sensitive to the effect of the geometrical parameters (distance between the molecular planes and rotation angle).

The selection of atomic net charges in the parameterization of Fraga's potential is a crucial task. A careful attention must be paid in testing that the electrostatic multipole moments of the molecule be quite well reproduced.

This work will be continued with the study of the growth mechanism of phthalocyanine clusters and β -crystalline phthalocyanine. The growth mechanism of the phthalocyanine clusters is simulated by building molecular stackings with a number of units from 1 to 10 in the β -crystal structure. The geometrical arrangement between the stackings has to be optimized (ref. 18).

ACKNOWLEDGEMENTS

One of us, F. T., acknowledges to the Conselleria de Cultura, Educació i Ciència de la Generalitat Valenciana for a Grant. This work has been partially supported by the project PS88-0112 of the D.G.I.C.Y.T. and the project 683.2/89 of the Universitat de València. This research has been made making use of the resources of the Servei d'Informàtica de la Universitat de València.

REFERENCES

- 1 E.S. Dodsworth, A.B.P. Lever, P. Seymour, and C.C. Leznoff, Intramolecular coupling in metal-free binuclear phthalocyanines, *J. Phys. Chem.*, 89 (1985) 5698.
- 2 E. Canadell and S. Alvarez, Theoretical study of the electrical behavior of one-dimensional metallophthalocyanines and related metallomacrocyclic compounds, *Inorg. Chem.*, 23 (1984) 573.
- 3 S. Fraga, A semiempirical formulation for the study of molecular interactions, *J. Comput. Chem.*, 3 (1982) 329.
- 4 S. Fraga, Molecular associations, *Comput. Phys. Commun.*, 29 (1983) 351.
- 5 E. Clementi, *Computational Aspects for Large Chemical Systems*, Springer-Verlag, Berlin, 1980.
- 6 J.C. Slater and J.G. Kirkwood, *Phys. Rev.*, 37 (1931) 682.
- 7 F. Torrens, J. Sánchez-Marín, E. Ortí, and I. Nebot-Gil, Incorporation of a dispersion energy term to Fraga's atom-atom pair intermolecular potential. Application to benzene, s-tetrazine, and their mixed dimers, *J. Chem. Soc., Perkin Trans. 2*, (1987) 943.
- 8 F. Torrens, A.M. Sánchez-de-Merás, and J. Sánchez-Marín, The use of "ab initio" net charges to improve Fraga's atom-atom pair potential for molecular association, *J. Mol. Struct. (Theochem)*, 166 (1988) 135.
- 9 F. Torrens, J. Sánchez-Marín, and F. Tomás, Intermolecular calculations on azine dimers, *J. Chem. Soc., Perkin Trans. 2* (submitted).
- 10 C. Douketis, G. Scoles, S. Marchetti, M. Zen, and A.J. Thakkar, Intermolecular forces via hybrid Hartree-Fock-SCF plus damped Dispersion (HFD) energy calculations. An improved spherical model, *J. Chem. Phys.*, 76 (1982) 3057.
- 11 R. Mason, G.A. Williams, and P.E. Fielding, Structural chemistry of phthalocyaninato-cobalt(II) and -manganese(II), *J. Chem. Soc., Dalton Trans.*, (1979) 676.
- 12 T.J. Marks, Electrically conductive metallomacrocyclic assemblies, *Science*, 227 (1985) 881.
- 13 A.M. Abeysekera, R. Grigg, J.F. Malone, T.J. King, and J.O. Morley, Chemical and theoretical studies of tautomerism in meso-aza- and meso-tetra-aza-porphyrins. Crystal structure of 5-aza-13,17-diethyl-2,3,7,8,12,18-hexamethylporphyrinatocobalt(II), *J. Chem. Soc., Perkin Trans. 2*, (1985) 395.
- 14 J. Pawliszyn, M.M. Szcześniak, and S. Scheiner, Interactions between aromatic systems: dimers of benzene and s-tetrazine, *J. Phys. Chem.*, 88 (1984) 1726.
- 15 W.C. Davidon, Variance algorithm for minimization, *Comput. J.*, 10 (1968) 406.
- 16 J.D. Head, M.C. Zerner, A Broyden-Fletcher-Goldfarb-Shanno optimization procedure for molecular geometries, *Chem. Phys. Lett.*, 122 (1985) 264.
- 17 F. Torrens, R. Montañana, and J. Sánchez-Marín, Vectorizing pair potential "AMYR" program for the study of molecular associations, in: J.L. Delhaye and E. Gelenbe (Eds.), *Proceedings of the European Symposium on High Performance Computing*, Elsevier, Amsterdam (accepted for publication).
- 18 F. Torrens, E. Ortí, and J. Sánchez-Marín, Electrically conductive phthalocyanine assemblies. Structural and non-integer oxidation number considerations, in: R.M. Metzger (Ed.), *Proceedings of the N.A.T.O. A.S.I. Lower-Dimensional Systems and Molecular Devices*, Plenum Press, New York (accepted for publication).

STRUCTURAL COMPARATIVE STUDY OF HYDROXY AND METHOXY FLAVONOIDS
USING MNDO, MINDO/3 AND AM1 PROGRAMS.

P.BONIFASSI¹, J.C.WALLET², E.M.GAYDOU²

¹ Centres de Calcul de l'I.U.T et de l'Université du Mans
Route de Laval, 72000, Le Mans, France

² Laboratoire de Phytochimie de Marseille, E.S.C.M
Faculté des Sciences de Saint Jérôme
Avenue Escadrille Normandie-Niemen, 13397, Marseille Cedex 13, France

SUMMARY

Structures of flavonoids given by experimental X-Ray data have been compared with calculations from various quantum chemistry methods.

INTRODUCTION

We have tried to rely geometric results obtained from X-ray data with calculated predictions using MNDO(ref. 1), MINDO/3(ref. 2) and a more recent AM1(ref. 3) methods for a few molecules listed in TABLE 1. We report here the preliminary results of this study.

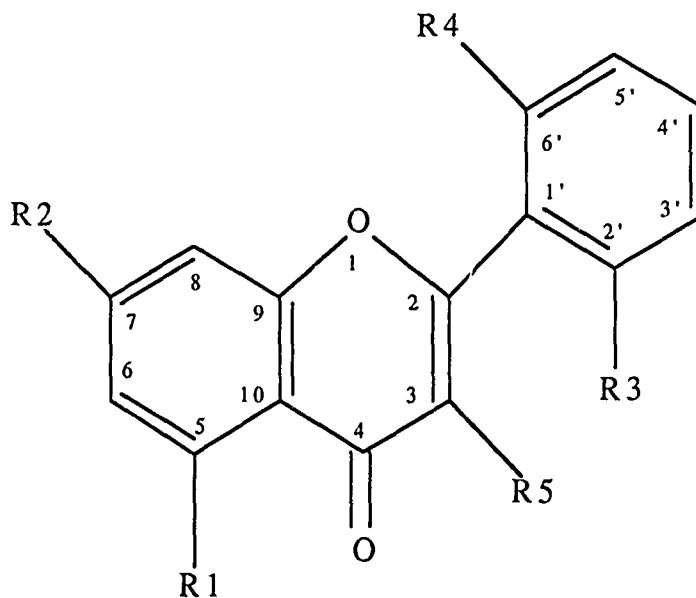


TABLE 1 : Entries of various molecules studied

	Flavonoid		R1	R2	R3	R4	R5
1	2'-methoxy	(ref. 4c)	H	H	OCH3	H	H
2	3-methoxy	(ref. 4a)	H	H	H	H	OCH3
3	2',6'-dimethoxy	(ref. 4b)	H	H	OCH3	OCH3	H
4	5-hydroxy-7-methoxy	(ref. 5)	OH	OCH3	H	H	H
5	3-hydroxy	(ref. 6)	H	H	H	H	OH

RESULTS AND DISCUSSIONS

Results in TABLE 2 and TABLE 3 are obtained when starting from experimental values with optimization of the torsion angle and intercycle distance and free rotation of methoxy groups.

TABLE 2 : Comparison of torsion angle by the different methods

	Experimental X-Ray	AM1	MNDO	MINDO/3
1	-2°.7	19°.7	-85°.4	121°.2
2	37°.2	37°.0	103°.8	—
3	70°.7	50°.6	102°.2	92°.6
4	21°.8	29°.2	79°	77°.6
5	5°.5	25°.1	78°	—

TABLE 3 : Comparison of distance of the intercycle bond by the different methods

	Experimental X-Ray	AM1	MNDO	MINDO/3
1	1.475A°	1.460A°	1.492A°	1.504A°
2	1.462A°	1.466A°	1.492A°	-
3	1.478A°	1.466A°	1.502A°	1.513A°
4	1.485A°	1.470A°	1.502A°	1.520A°
5	1.474A°	1.458A°	-	-

MNDO and MINDO/3 methods overestimate the stability of the perpendicular conformation even without substitution on positions 2' and 6' (entry 4) .

In each case, higher values are obtained for torsion angles and intercycle bonds .

Using AM1 program conducts to satisfactory agreement for bond lengths since errors are less than 0.02A° . With regard to torsion angles we get the same value for the 3-methoxy flavanol and a very small difference for the 5-hydroxy-7-methoxyflavone. There is a difference of 20° for the three other cases (entries 1, 3 and 5) .

In crystalline state, entries 3 and 5 involve intermolecular hydrogen bonds and entry 1 two intra molecular hydrogen bonds which in the later case constrain the two parts of the molecule to be planar . Such hydrogen bonds are a possible explanation to the observed deviations .

A study of reaction profile shows very little variation of heat of formation (0.1 Kcal/mole) when angle of intercycle torsion varied from 40° to 140° , the intercycle distance being optimized for each angle value .

Perhaps, the observed deviations have no physical meaning for a gaseous molecule, because there is free rotation of the phenyl ring between 40° and 140°.

No doubt, a new study by MNDO with new parameters of J.J.P.Stewart (ref. 8 - 9) would give similar results as by AM1.

CONCLUSION

Compared to MNDO and MINDO/3, AM1 gives better results in the present study , the nuclear repulsions being overestimated in MNDO and in MINDO/3 .

Discrepancy between X-Ray results and calculated geometries using AM1 may be due to the fact that we consider simple molecules which do not take into account near neighbouring interactions(Ref. 9),especially inter molecular hydrogen bonds.

REFERENCES

- 1 M.J.S.Dewar and W.Thiel J. Am. Chem. Soc. (1977) 99 , 4899, 4907.
- 2 R.C.Bingham, M. J. S. Dewar and V. D. H. Lo J. Am. Chem. Soc (1975) 97, 1285, 1302, 1307.
- 3 M. J. S. Dewar, F. G. Zoebisch, E. F. Healy and J. J. P. Stewart J. Am. Chem. Soc. (1985) 107, 3902.
- 4 a) J. C. Wallet, E. M. Gaydou, A. Fadlane and A. Baldy Acta Cryst. (1988) C44, 357.
b) J. C. Wallet, E. M. Gaydou and A. Baldy Acta Cryst. (1989) C45, 512.
c) J. C. Wallet, E. M. Gaydou, J. Jaud and A. Baldy Acta Cryst. admitted for publication.
- 5 M. Shoja Acta Cryst. (1989) C45, 828.
- 6 M. C. Etter, Z. Urbanczyk- Lipkowska, S. Baer and P. F. Barbara. J. Mol. Struct.(1986) 144,155.
- 7 J. J. P. Stewart J. Comput. Chem. (1989) 10, 2, 209.
- 8 J. J. P. Stewart J. Comput. Chem. (1989) 10, 2, 221.
- 9 L. K. Vinson and J. J. Dannenberg J. Am. Chem. Soc. (1989) 111, 2777.

THEORETICAL STUDY OF THE STRUCTURE OF THE GLUTATHIONE - HYDROGEN PEROXIDE COMPLEX

J. BERGES¹, J. CAILLET¹, J. LANGLET¹, Z. ABEDINZADEH²,
M. GARDES-ALBERT²

¹Dynamique des Interactions Moléculaires, Université Pierre et Marie Curie, 4 Place Jussieu, 75005 Paris (France)

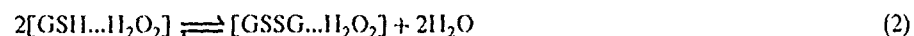
²Laboratoire de Chimie Physique, Université René Descartes, 45 Rue des Saints Pères, 75006 Paris (France)

SUMMARY. The possibility of intermolecular complexes between Glutathione and hydrogen peroxide has been investigated in the framework of the SIBFA method. As a first step, a study of some local minima of Glutathione in both an isolated state and water has been performed. The results are compared with available experimental data.

INTRODUCTION

Glutathione (GSH = L-γ-glutamyl-L-cysteinyl-L-glycine) is the major non protein thiol compound present in cells. Nucleophilic and reductant properties of the thiol group confers to Glutathione a variety of roles, one of the most important being the reduction of hydrogen peroxide H₂O₂ by GSH mediated by glutathione peroxidase (ref. 1).

Recently, the reaction of H₂O₂ with GSH in absence of enzyme, has been studied in vitro conditions by Abedinzadeh et al. (ref. 2). They have shown the initial fast formation of a peroxide or chelate between GSH and H₂O₂ (reaction (1)) followed by the oxidation of [GSH...H₂O₂] into [GSSG...H₂O₂](reaction (2))



From a theoretical point of view, it seemed interesting to test the possibility of such an H-bonded complex formation. The present work is a preliminary study of this problem involving both intra and intermolecular calculations without and with solvent interactions.

Since experimental work has been carried out in aqueous medium buffered at pH 7.4, we were only interested by the negative ion [GSH]⁻ (see Fig.1) which represents the state of dissociation of Glutathione at this pH value according to both H-NMR (ref. 3) and ¹³C-NMR (ref. 4) studies.

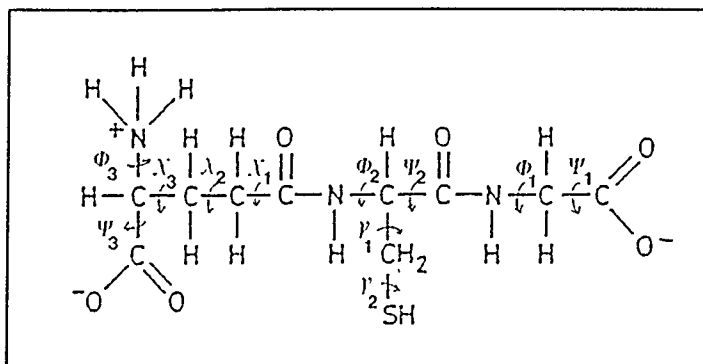


Figure 1. Negative Ion Glutathione at pH 7.4
Definition of the degrees of rotational freedom

Our work may be divided into two parts:

- A study of the conformation of $[\text{GSII}]^-$ both in an isolated state and in water.
- A study of intermolecular interactions between H_2O_2 and $[\text{GSII}]^-$, taking into account the eventual conformation change of both $[\text{GSII}]^-$ and H_2O_2 . The three different structures of $[\text{GSII} \cdots \text{H}_2\text{O}_2]$ proposed by Abedinzadeh et al. (ref. 2) have been investigated.

As emphasized in section II_A, we have not undertaken a complete exploration of the entire $[\text{GSII}]^-$ conformational space, in fact in this preliminary work, our purpose was to study the behaviour of a limited number of local minimum conformations with regards to $[\text{GSII} \cdots \text{H}_2\text{O}_2]$ H-bonded complex formation.

I METHOD

Both intra and intermolecular energies have been calculated simultaneously in the framework of the SIBFA method (*Sum of Interactions Between Fragments computed Ab-initio*). For more details concerning this method, see ref. 5. In therein work, the charge distribution of each fragment has been obtained from a multipolar expansion (ref. 6) of ab-initio SCF wave function calculated within an adapted minimal basis set (ref. 7).

As an evaluation of the solvent effect, we have only taken into consideration 'Hydration water' molecules, i.e. the ones which are very close to the solute and thus interact very strongly with it. In order to estimate the 'Hydration energy' (ΔE_{hydra}), it may be supposed that each water-solute interaction (E_{w-s}) replaces a water-water interaction (E_{w-w}), N_w being the number of 'Hydration water' molecules; we get:

$$\Delta E_{\text{Hydra}} = \sum_{w=1}^{N_w} E_{w-s} - N_w E_{w-w}$$

We have used the value of E_{w-w} of 7.4 Kcal/mole calculated within the SIBFA method. We are conscious that ΔE_{H_2O} only represents a part (but an important one) of the total solvation energy in water, but such a study should give an eventual insight into possible intramolecular conformational change due to these strong water-solute interactions.

II CONFORMATION OF $[GSII]^-$

In spite of its wide biochemical interest, experimental and theoretical data on the conformation of Glutathione are rather scarce:

- A crystal structure determination has been reported for the neutral form GSII (ref. 8).
- Among the few NMR studies carried on Glutathione, only one (ref. 3) was concerned by the geometrical arrangement in water solution at different pH values.
- PCILO calculations (ref. 9) have been performed for both GSII and $[GSII]^-$ in an isolated state.

All these results are somewhat sparse and furthermore they are not consistent. But such an outcome principally proceeds on one hand from the flexibility of Glutathione (even when keeping the two peptide links planar with NH and CO groups in a trans position, it still remains eleven degrees of rotational freedom (see Fig. 1)) and on the other hand from the presence of the three functional groups (two acid plus one amino groups), the dissociation of which strongly depends on the pH value. Consequently as emphasized by (ref. 3,9), the conformational space should include a great deal of local minima, the relative importance of which is related to change in the functional groups and to environmental effects.

A *Isolated state*

We have used bond lengths and angles determined in the crystal structure (ref. 8). Our results proceed from a simultaneous variation of the eleven torsional angles (defined in Fig. 1) through an optimization process carried out with two initial guesses:

- The S-shaped 'open' conformation (denoted S) observed in the GSII crystal structure (Fig. 2a).
- A local energy minimum conformation (denoted S') obtained from a sub-map $E = f(\Psi_2, \Phi_2)$, all other torsional angles being frozen at the values of (ref. 8). This choice reflects the fact that Ψ_2 and Φ_2 dihedral angles are important with regards to the relative position of the two peptide links and thus for the geometrical arrangement of the central part of the molecule, i.e. the one probably involved (following ref. 2) in the $[GSII \cdots H_2O]$ interaction. Optimization of S and S' have respectively led to Opt₁ (Fig. 2b) and Opt₂ (Fig. 2c) conformations.

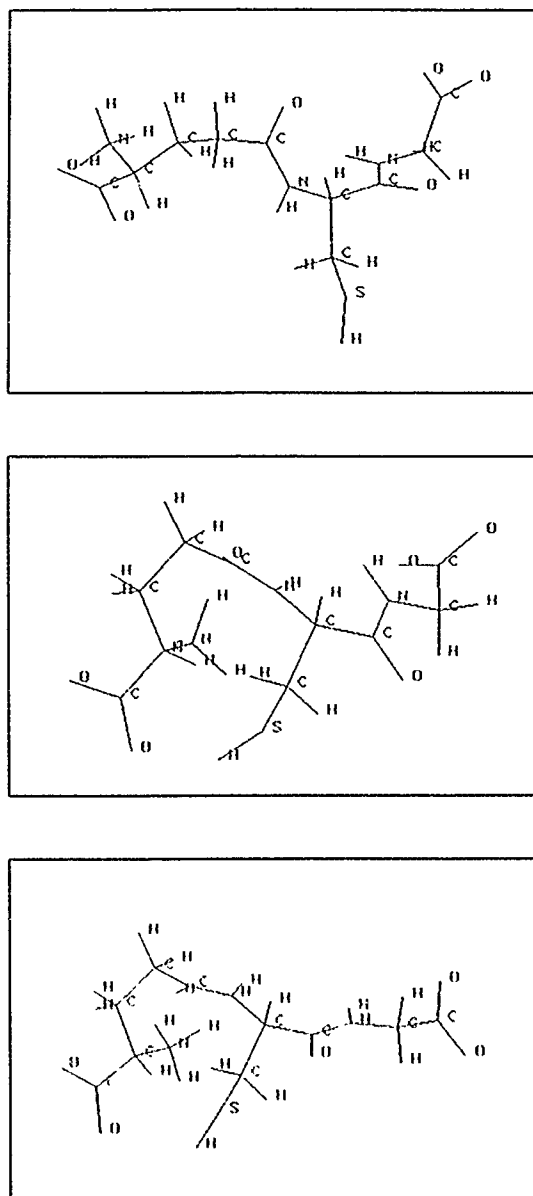


Figure 2. Glutathione conformations in the isolated state:
a) from RX data (S). b) Calculated Opt₁. c) Calculated Opt₂.

The intramolecular energy differences between S and both Opt₁ and Opt₂ may be found in Table 1. The values of the main dihedral angles calculated in S, Opt₁ and Opt₂ are listed in Table 2.

Confor- mations	ΔE_{Repul}	ΔE_{Disp}	ΔE_{El}	ΔE_{Pol}	ΔE_{Tot}
Opt ₁	14.9	-10.1	-55.8	2.6	-54.9
Opt ₂	9.3	-7.4	-53.9	2.3	-56.1
Opt ₃	8.3	-6.7	-46.1	2.3	-40.6

Table 1. Variation of intramolecular energy ΔE_{Tot} (and its components).
Subscripts: *Repul*, *Disp*, *El*, *Pol*, for Repulsion, Dispersion, Electrostatic
and Polarisation respectively. (All values in Kcal/mole).

Confor- mations	Φ_1	Ψ_1	Φ_2	Ψ_2	χ_2	χ_3
S	275.1	190.0	269.1	356.5	219.0	289.0
Opt ₂	172.2	119.1	197.8	22.0	305.7	297.6
Opt ₃	193.2	123.8	178.0	126.1	300.0	314.7
Opt ₄	196.4	121.6	193.7	42.0	291.4	280.4

Table 2. Values of some dihedral angles defining different Glutathione conformations.
The values calculated for the five other angles are : $\gamma_1 \approx 60^\circ$, $\gamma_2 \approx 180^\circ$,
 $\chi_1 \approx 90^\circ$, $\Phi_3 \approx 60^\circ$, $\Psi_3 \approx 30^\circ$ (291° in S conformation). (All values in degrees).

From an energetical point of view, our results clearly show :

- A drastic instability of S conformation with regards to both Opt₁ and Opt₂. The energy difference we have calculated between S and Opt₁ or Opt₂ : $\Delta E \approx 55$ Kcal/mole mainly proceeds from the electrostatic component of the intramolecular energy.
- An almost equal stability between both Opt₁ and Opt₂ : The 1.2 Kcal/mole energy difference we have obtained is not very significant.

From a geometrical point of view, in agreement with PCILO results (ref. 9), the optimization of S (leading to Opt₁) does not change significantly the central part of the molecule (see Table 2); we have just noticed a modification of the value of the angle α between the planes containing the two peptide linkages ($92^\circ 0$ and $43^\circ 6$ respectively) while the geometrical arrangement of the glycyl residue and of the glutamyl chain are rather different, effectively.

- A displacement of the glycyl carboxyl group out of the plane $O_3N_1C_2$ is observed in S conformation ($\Phi_1 = 275^\circ$) but not in Opt_1 ($\Phi_1 = 172^\circ$), but in this latter it may be noticed a torsion of glycyl carboxylate oxygen around the C-C bond ($\Psi_1 = 119^\circ$), thus in contrast with the PCILO results the glycyl residue does not adopt a C_s conformation.

- Following the nomenclature of Fujiwara et al. (ref. 3) the glutamyl chain appears to be a distorted 'open' rotamer I-3 in S conformation and a closed rotamer II-3 in Opt_1 . An insight to Fig. 2a,b clearly shows that neither S nor Opt_1 geometries involves any H-bond between the NH_3^+ or COO^- glutamyl groups and the peptide backbone, but in Opt_1 a weak H-bond ($d_{O...H} = 2.4\text{\AA}$) occurs between carboxylate group and cysteic SH group. Once again our local minimum is different from PCILO one which indicates a preference for structures with glutamyl carboxylate in interaction with the peptide link. This will be discussed later on.

- Opt_1 and Opt_2 mainly differ by the value of Ψ_2 angle ($83^\circ 5$ and $187^\circ 5$ respectively), consequently, in Opt_2 the two peptide links are in a nearly trans-position, the α angle (previously defined) has significantly varied: $224^\circ 9$ instead of $43^\circ 6$. Furthermore Fig. 2c indicates a very weak interaction ($d_{H...S} = 2.95\text{\AA}$) between NH_3^+ glutamyl group and SH cysteic group.

B In water.

As a first step, we have taken into account ten 'Hydration water' molecules: the intramolecular conformation of S, Opt_1 and Opt_2 does not change significantly and now hydrated Opt_1 is stabilized with regards to both hydrated S and Opt_2 conformations; furthermore Opt_2 appears to be saturated (in sense of 'Hydration water').

Our process was carried on with Opt_1 : the saturation occurs with eighteen 'Hydration water' molecules; as a result of simultaneous inter and intramolecular optimization processes we have obtained Opt_3 (see Fig. 3) which is practically similar to Opt_1 from a geometrical point of view (the eleven dihedral angles values do not differ by more than 20°). From an energetical point of view it may be noticed that the 14.3 Kcal/Mole loss in intramolecular energy is balanced by a 127 Kcal/Mole hydration gain. Fig. 3 also gives an illustration of the position of the eighteen water molecules interacting with Opt_3 . As appears in this figure, one water molecule form an H-bond bridge between NH_3^+ and COO^- groups of the glutamyl fragment. We have observed a similar situation with ab-initio calculations (within 6-31G** basis set) of glycine molecule which has the same topology than the one existing in the $COO^-CH_2NH_3^+$ group of the glutamyl part of [GSH].

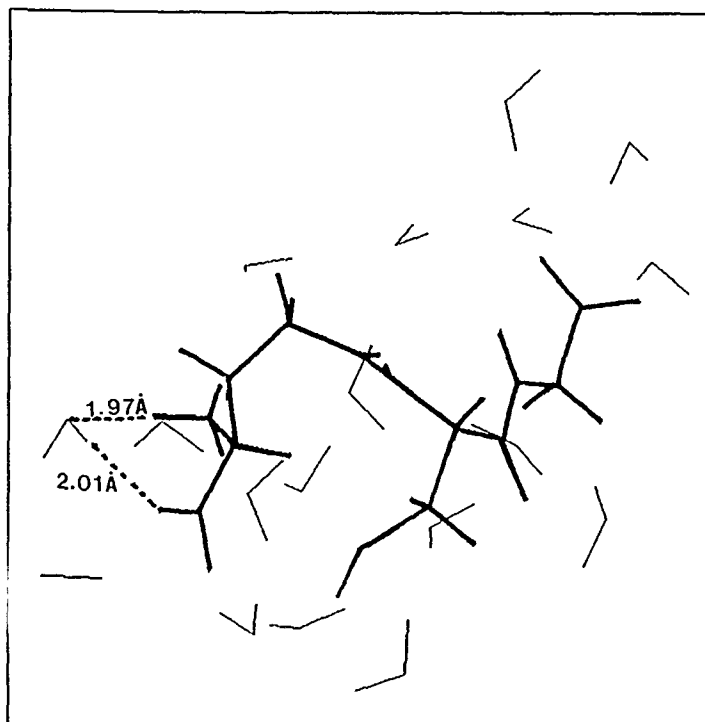


Figure 3. Glutathione in Opt_3 conformation with 'Hydration water' molecules. GSII is represented by heavy line. Some molecular distances between GSII and water are represented by dashed lines.

III $[\text{GSII}]^- \cdots \text{H}_2\text{O}_2$ COMPLEXES

As an initial guess, we have taken the $[\text{GSII}]^-$ conformation (Opt_3) optimized in water and the experimental skew form of H_2O_2 ($\tau = 120^\circ$). Not any optimization process has been performed on the isolated state of H_2O_2 since it is well known that available results on H_2O_2 are only obtained in the framework of ab-initio calculation fulfilling at least two criteria: 1) The basis set employed has to be augmented by polarization functions. 2) All geometrical parameters have to be optimized for all values of τ to be considered (ref. 10). In this preliminary work, we have studied the three $[\text{GSII}]^- \cdots \text{H}_2\text{O}_2$ complexes suggested by Abedinzadeh et al. (ref. 2) namely the ones involving interactions between H_2O_2 and 1) both CO groups: $[(\text{Opt}_3 \cdots \text{H}_2\text{O}_2)_{\text{CO,CO}}]$, 2) both NH group belonging to the two peptide links: $[(\text{Opt}_3 \cdots \text{H}_2\text{O}_2)_{\text{NH,NH}}]$, 3) glycyl NH and glutamyl CO group: $[(\text{Opt}_3 \cdots \text{H}_2\text{O}_2)_{\text{CO,NH}}]$. As a first step, only intermolecular and H_2O_2 intramolecular energies have been optimized. Table 3 indicates that $[(\text{Opt}_3 \cdots \text{H}_2\text{O}_2)_{\text{NH,NH}}]$ seems to be the most stable complex.

	E_{Repul}	E_{Disp}	E_{El}	E_{Pol}	E_{Chl}	E_{Tot}
$(Opt_3 - H_2O_2)_{CO,NH}$	6.0	-5.8	-7.1	-0.3	-0.8	-8.0
$(Opt_3 - H_2O_2)_{NH,NH}$	8.5	-7.3	-8.2	-0.8	-3.7	-11.6
$(Opt_3 - H_2O_2)_{CO,CO}$	9.1	-7.4	-8.2	-0.3	-0.8	-7.5
$(Opt_4 - H_2O_2)_{CO,NH}$	8.7	-7.6	-11.6	-0.5	-1.9	-12.9

Table 3. Total intermolecular interaction energy (and its different components) for some $GSII \cdots H_2O_2$ complexes.

Subscripts: *Repul*, *Disp*, *El*, *Pol* have the same meaning as in the Table 1. E_{Chl} stands for the charge-transfer component. (All values in Kcal/mole).

An insight to Fig. 4 shows that, in fact, the interactions between CO or NH peptides groups should be very weak since the shortest distances thus calculated lies between 2.44Å and 2.74Å. In fact, the stability of $[(Opt_3 \cdots H_2O_2)]$ complex mainly proceeds from a strong H-bond interaction between H_2O_2 and glycyl COO^- group ($d_{H \cdots O} = 1.98\text{Å}$). It has been noticed that the value of the angle τ (defining the conformation of H_2O_2) does not deviate significantly from 120° , the value obtained for the minimal conformation.

As a second step, possible changes of both $[GSII]^-$ and H_2O_2 geometries have been taken into account in the optimization process: this does not affect markedly $[(Opt_3 \cdots H_2O_2)]_{NH,NH}$ and $[(Opt_3 \cdots H_2O_2)]_{CO,CO}$ complexes, but stabilizes the one involving glycyl NH and glutamyl CO groups: $[(Opt_4 \cdots H_2O_2)]_{CO,NH}$ is thus obtained (see Table 3) strikingly, as shown by Fig. 4b, the optimization of $[GSII]$ geometry brings H_2O_2 nearer to glycyl COO^- group ($d_{H \cdots O} = 1.94\text{Å}$), leading to a strong H-bond interaction. The geometrical parameters defining Opt_4 conformation are practically identical to the ones calculated for Opt_3 (variation of dihedral angles is less than 20°), and the intramolecular energy remains almost unchanged (variation of intramolecular energy is less than 2Kcal/mole).

In this preliminary work, we have not taken into account the solvent effect, thus we cannot decide that one of the four complexes we have obtained is the most stable one; we may only conclude that effectively intermolecular complexes between H_2O_2 and $[GSII]^-$ are possible. Furthermore, at the light of our results, CO and NH peptide group are not the most favourable interacting sites; we have shown up the involvement of glycyl COO^- group in the complexation phenomenon.

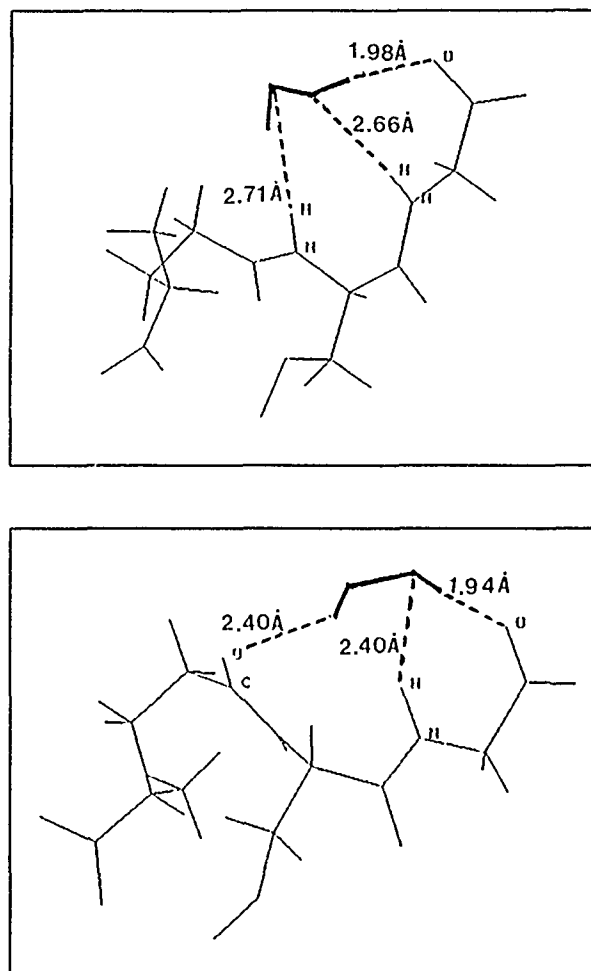


Figure 4. Glutathione - Peroxide hydrogen Complexes
 a) $[\text{Opt}_3 - \text{H}_2\text{O}_2]_{\text{Ni}(\text{NH}_4)}$ b) $[\text{Opt}_4 - \text{H}_2\text{O}_2]_{\text{Co}(\text{NH}_4)}$
 H_2O_2 is represented by a heavy line. Shortest intermolecular distances are represented by dashed lines.

CONCLUSION

In the light of the results of the calculations reported above, three main conclusions may be drawn-up:

1) The geometrical arrangement observed in the crystal structure of Glutathione does not represent a local minimum either in an isolated state or in water solvent. It must be emphasized that this determination only concerns the neutral form of Glutathione and not its negative ion which may adopt a different conformation. This hypothesis is supported by PC11.0 calculations which have shown that changes in the functional groups of glutamic acid have an important influence on the preferred conformation of

the molecule. Furthermore, the molecular geometry found in GSII crystal should mainly proceed from environmental effects, since, in this structure, the GSII molecules are held together by a three dimensional network of hydrogen bonds involving carboxyl and amino terminal groups and peptide nitrogen and oxygen atoms.

2) In the conformational space, besides closed (gauche) conformation involving intramolecular II-bonds between terminal glutamyl ionized groups and atoms belonging to the two peptide links (like the III-I rotamer found in PCILO calculations) it exists some closed conformations in which neither CO nor NH peptide groups are involved into internal II-bonds. This result is not inconsistent with NMR conclusions which indicate that, at the physiological pH the gauche population (60%) may include both PCILO and our conformations.

3) Intermolecular complexes between GSII and H_2O_2 are possible: this result is not without interest from a biological point of view, since it may be conceived that even in vivo, in absence of Glutathione peroxydase, Glutathione may mask H_2O_2 by complexing it.

We are conscious that this work only represents a first step as concerns the answer of our problem. Our results can be considered only as qualitative ones. We have especially studied the possibility of some conformations of GSII to induce some intermolecular complexes with H_2O_2 . Obtaining quantitative results (which should led to association constant for $[\text{GSII}] \dots \text{H}_2\text{O}_2$ complex) requires the exploration of a more complete $[\text{GSII}]$ conformational space, the taking into account of the whole solvent and the study of complex configurations which may be different than the ones proposed by Abedinzadeh et al (ref. 2). These calculations have been undertaken at the present time.

The authors acknowledge financial support of these studies by the "Groupement Scientifique IBM-CNRS: Modélisation moléculaire".

REFERENCES

1. B. Halliwell, J.M.C. Gutteridge, Protection against oxygen radicals in biological systems in: Free radicals in Biology and Medicine, Oxford University Press (1986) pp.67-136.
2. Z. Abedinzadeh, M. Gardès-Albert, C. Ferradini, Kinetic Study of The Oxidation Mechanism of Glutathione by Hydrogen Peroxide in Neutral Aqueous Medium, Can. J. Chemistry, 67 (1989) 1247-1255.
3. S. Fujiwara, G. Formicka-Kozłowska and H. Kozłowski, Conformational Study of Glutathione by NMR, Bull. Chem. Soc. Japan, 50 (1977) 3131
4. T. N. Huckerby, A.J. Tudor and J. G. Dawber, Acid-Base Studies of Glutathione (L- γ -Glutamyl-L-cysteinyl-L-glycine) by One- and Two-dimensional Nuclear Magnetic Resonance Spectroscopy, Perkins Trans. II (1985) 759

5. N. Gresh, P. Claverie and A. Pullman, Theoretical studies of molecular conformation. Derivation of an additive procedure for the computation of intramolecular interaction energies. Comparison with ab-initio SCF computations. *Theoret. chim. Acta* 66 (1984) 1-20
6. F. Vigné-Maeder, P. Claverie, The exact multicenter multipolar part of a molecular charge distribution and its simplified representations, *J. Chem. Phys.* 88 (1988) 4934-4948
7. H. Berthod and A. Pullman, *J. Comput. Chem.* 87 (1987) 2- Molecular potential, cation binding, and hydration properties of the carboxylate anion. Ab initio studies with an extended polarized basis set, *J. Comput. Chem.* 2 (1987) 87-95
8. W.B. Wright, The crystal structure of Glutathione, *Acta Cryst.* 11 (1958) 632-642
9. P.R. Laurence, C. Thomson, A Theoretical Investigation of the Preferred Conformations of Glutathione and Its Constituent Amino Acid Residues. *Theoret. Chim. Acta*, 57 (1980) 25-41.
10. D. Cremer, General and theoretical aspects of the peroxide group in: S. Patai (Ed.), *The Chemistry of Functional Groups, Peroxides*, John Wiley and Sons Ltd, New York, 1983, pp. 2-

STRUCTURE AND DYNAMICS OF WATER AND IONIC SOLUTIONS NEAR
BIOMEMBRANE SURFACES FROM MOLECULAR DYNAMICS SIMULATIONS

M. SCHLENKRICH¹, K. NICKLAS¹, J. BÖCKER¹, P. BOPP² and
J. BRICKMANN¹

¹Institut für Physikalische Chemie
Technische Hochschule Darmstadt
Petersenstrasse 20, 6100 DARMSTADT (Federal Republic of Germany)

²Institut für Physikalische Chemie
Rheinisch Westfälische Technische Hochschule Aachen
Templergraben 59, 5100 AACHEN (Federal Republic of Germany)

SUMMARY

The interface between a membrane, modelled by an ensemble of COO⁻ groups with rotational and translational degrees of freedom, and liquid water with solvated ions is studied by molecular dynamics (MD) computer simulations. The charged membrane leads to a layering of the liquid. Several water layers can be distinguished with structural and dynamic properties very different from those found in the bulk phase.

INTRODUCTION

A detailed knowledge of the microscopic structure and dynamics in the interfacial region between water, or aqueous solutions, on one side, and a biomembrane on the other, is an essential prerequisite for the understanding of many elementary biological processes [1].

In the present work, we employ the molecular dynamics (MD) computer simulation method in a first attempt to study the typical phenomena which may occur at such an interface between pure water or a solution and a membrane. Presently we do not attempt to give a quantitative description, but rather to determine typical features and trends.

THE MODEL AND DETAILS OF THE SIMULATION

The system studied consisted of 737 TIP4P water molecules and of 60 COO⁻ groups representing the membrane surfaces, one on each side of a water lamina (see fig.1). The density of the headgroups

was varied between 0.042 \AA^{-2} and 0.05 \AA^{-2} . Na^+ ions were added to the system until charge neutralization was reached.

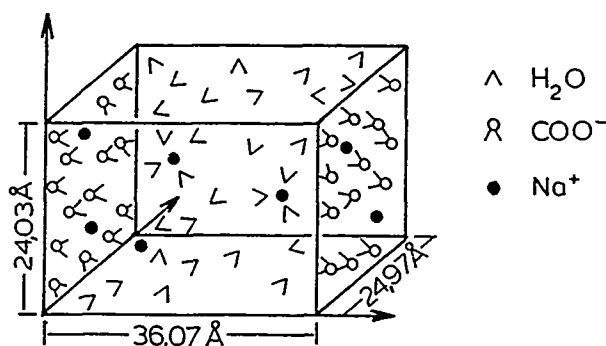


Fig.1 Model system (schematically)

The total potential energy of the system is partitioned into the following contributions:

$$V_{\text{tot}} = V_{\text{W-W}} + V_{\text{W-S}} + V_{\text{S-S}} + V_{\text{W-I}} + V_{\text{S-I}} + V_{\text{I-I}} \quad (1)$$

where W stands for a water molecule, S for a headgroup in the membrane (COO^- groups) and I for a Na^+ ion. The TIP4P model potential [3] is used for the intermolecular water-water interaction $V_{\text{W-W}}$. The other interaction potentials were developed accordingly along the same lines as sums of Coulomb and Lennard Jones (12,6) site-site pair potentials, using a model by Jorgensen and Gao [4]. A model potential due to Bounds [5] was used for $V_{\text{W-I}}$. The simulations are carried out in the usual fashion at constant number of particles, volume and total energy (NVE-MD). Full details of the model-potentials and of the simulations are given elsewhere [6].

STRUCTURAL RESULTS OF THE INTERFACIAL SYSTEM

The Structure of Aqueous Phase

A strong influence of the charged groups of the membrane on the adjacent liquid is to be expected from the work on ionic hydration [7] and on charged surfaces [8]. In the interfacial region three structuring effects, namely from the bulk phase, from

the membrane, and from the ionic hydration, are in competition, leading to interesting cooperative effects (fig. 2).

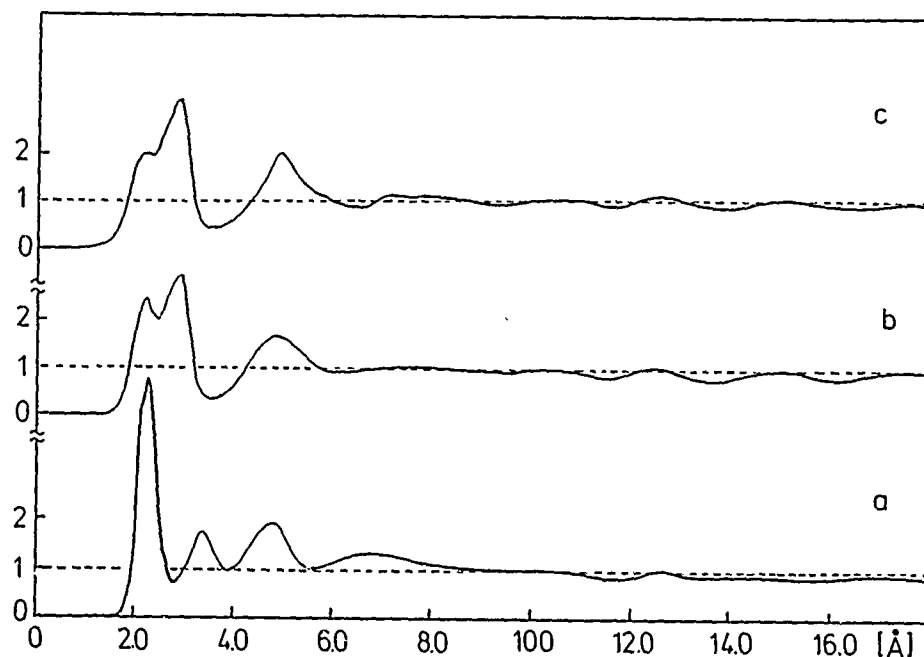


Fig.2 Density profiles for the water molecule oxygens as a function of their distance from the plane of the carbon atom for pure water (a), 35 ions (b) and 60 ions (c).

It is seen that the presence of the ions leads to a decrease of the strong density oscillations observed in pure water. Due to the adsorption of ions in the immediate vicinity of the membrane (vide infra), the first two layers are collapsed into one. The overall increase in water density in the vicinity of the membrane, by a factor of about two compared to pure water, remains roughly the same in all systems. Such density increases have been found experimentally by Joosten [9].

Figure 3 shows the density profiles for the ions. These were added slowly to the pure water system by replacing water molecules near the center of the lamina. About 4 to 8 ions were added at a time, and the system was then allowed to reequilibrate.

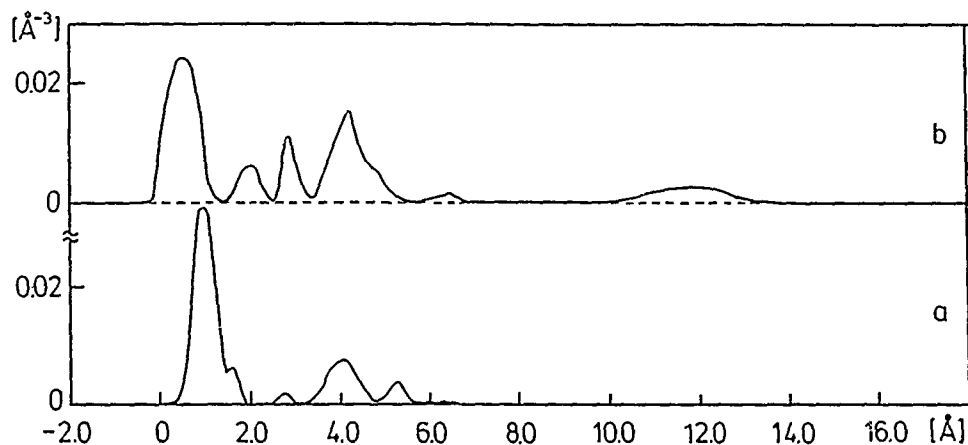


Fig.3 Density profiles for sodium ions as functions of the distance from the membrane surface.
35 ions (a), 60 ions (b).

The interesting feature is the saturation of the first layer of adsorbed ions at about 12 ions vs. 30 COO^- groups. This corresponds to a particularly favorable geometric arrangement of the ions with respect to the membrane groups. Further addition of ions lead to the formation of a weakly pronounced second layer and finally of a diffuse distribution throughout the lamina.

Density fluctuations as seen in figure 2 are an indication of a layering of the water molecules parallel to the membranes and suggest the interpretation of this phenomenon as an adsorption. We shall make use of this concept of 'adsorbed' layers of water in order to further characterize the structure of the water adjacent to the membranes. Between these adsorbed layers and the bulk phase a structure broken region can be identified [6]. Figure 4 shows distribution functions of the oxygen distances within the first layer of adsorbed water with and without ions, and for the "bulk" phase (central part of the water lamina).

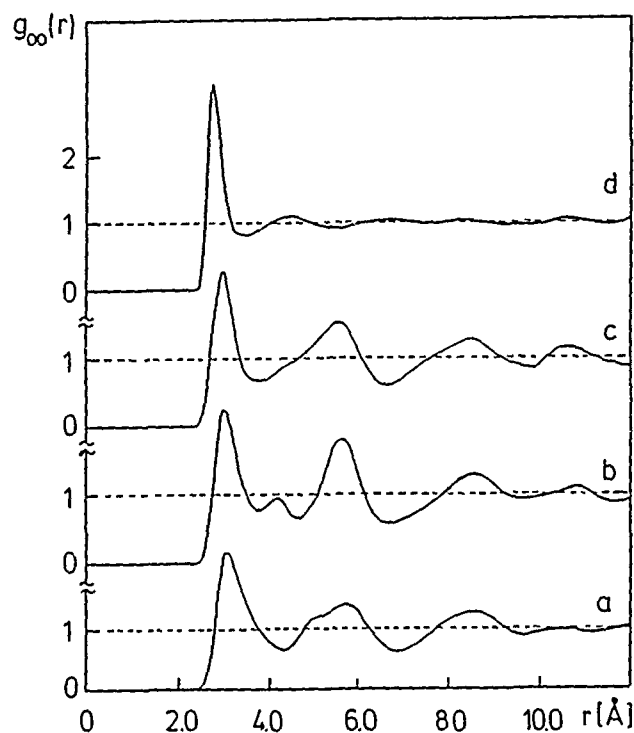


Fig.4 Oxygen-oxygen pair distribution functions
 pure water, first layer (a), 35 ions, first layer (b),
 60 ions, first layer (c), 60 ions, bulk phase (d).

In the pure water system, the ordering of the water in the first layer is almost completely determined by the structure of the membrane (a). Due to the hydration of ions, with a typical Na^+ -water distance of 2.3 Å, and to the influence of the ions on the membrane itself, this structure is modified in the presence of ions (b,c). The structure broken transition region found between the water layers and the bulk water in the presence of ions is much less pronounced than in the pure water case.

THE DYNAMICS OF THE MEMBRANE-LIQUID INTERFACE

There are substantial differences between the dynamical behavior of bare model membranes and such in contact with water or ionic solution [6]. Fluid like motions with very large amplitudes and occasional exchanges take place in the bare membrane. These motions are quite efficiently hindered by the adsorbed water molecules and also by the ions.

Detailed information about the dynamics can be obtained from the velocity autocorrelation function

$$c_{VV}(t) = 1/T \int_0^T \langle v_i(\tau+t) \cdot v_i(\tau) \rangle d\tau \quad (2)$$

and its Fourier cosine transform (spectral density, fig.5)

$$\hat{c}_{VV}(\omega) = \int_0^\infty c_{VV}(t) \cdot \cos(\omega t) \cdot dt \quad (3)$$

The zero frequency term of $\hat{c}_{VV}(\omega)$ is proportional to the self-diffusion coefficient. The averages in equation (2) (pointed brackets) have been carried out over water layers parallel to the membrane.

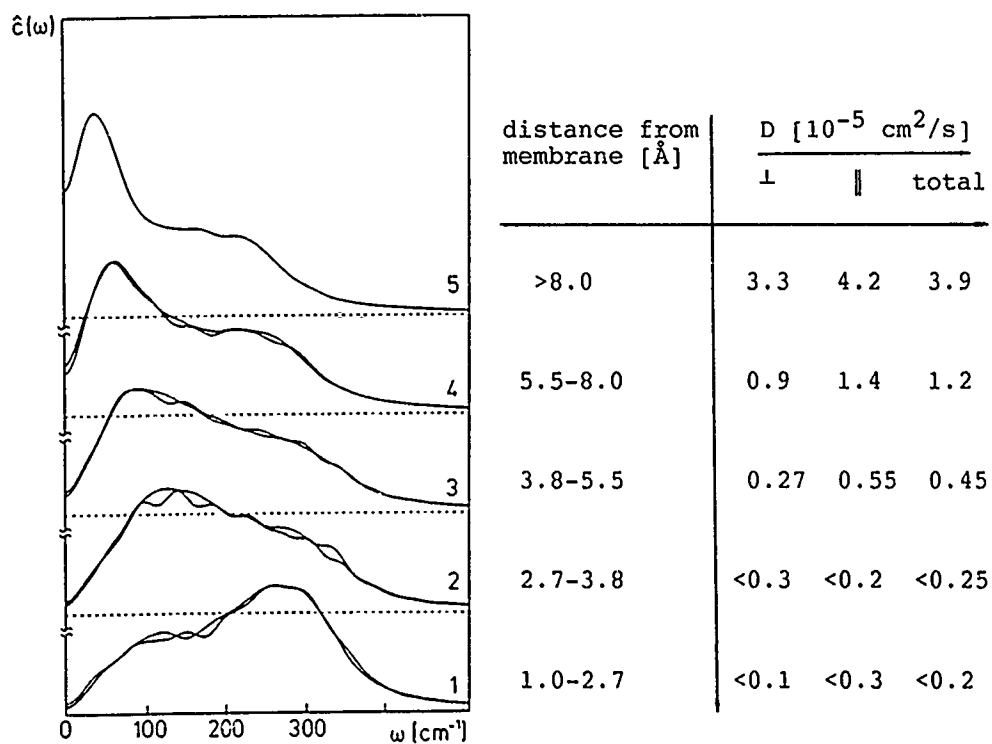


Fig.5 Spectral densities obtained for the hindered translations of the water molecules in various layers. The self-diffusion coefficients, determined separately for motions perpendicular and parallel to the surface, are also given.

The immobilization of the water by the membrane leads to a decrease of the self diffusion by a factor of about 10. A similar trend has been found in simulations of smectic liquid crystals [10]. Even in the center of the lamina, where the structure strongly resembles the one of pure water, the self diffusion remains somewhat anisotropic. Except for this feature, the spectral density with its characteristic peak at 50 cm^{-1} and the pronounced shoulder around 200 cm^{-1} is very similar to the one for pure water [11]. Moving closer to the membrane, it is continuously altered and shifted to higher frequencies. A detailed analysis reveals that the peak at lower frequency in the adsorbed layer is due mostly to motions of the water molecules parallel to the membrane surface. This peak has been found to be sensitive to the density of the headgroups in the membrane.

CONCLUSION

We have investigated the interface aqueous solution / charged membrane with respect to the static and dynamic properties of the liquid and of the membrane. It was attempted to make the system sufficiently large to include the complete transition region between bulk water and water adsorbed to the membrane. The density of the water is increased by a factor of about 2 next to the membrane. The overall structural influence of the membrane extends about 8 to 9 Å into the liquid. This range and the details of the ordering depend on the surface density of the membrane and on the ionic concentration. The single molecule dynamics of the water reveals a longer range of influence. For the hindered translation, the transition between ordered layers and bulk is continuous, in contrast with the finding for the structure.

ACKNOWLEDGEMENT

A grant of computer time by the Höchstleistungsrechenzentrum (HLRZ) at the Kernforschungsanlage Jülich is acknowledged. Financial support from the Deutsche Forschungsgemeinschaft and from the Fonds der Chemischen Industrie, Frankfurt is also acknowledged.

REFERENCES

- [1] A.Skerra and J.Brickmann, *Biophys.J.* 51, 969 (1987), *ibid.* 51, 977, (1987), and references therein, M.Schlenkrich, P.Bopp, M.Knoblauch, A.Skerra and J.Brickmann, "Advances in Biotechnology of Membrane Ion Transport", edited by P.L.Jorgensen and A.Verna, Raven Press, New York 1988.
- [3] W.L.Jorgensen, J.Chandrasekhar, J.D.Madura, R.W.Impey and M.L.Klein, *J.Chem.Phys.* 79, 926 (1983).
- [4] W.L.Jorgensen and J.Gao, *J.Phys.Chem.* 90, 2174 (1986).
- [5] D.G.Bounds, *Mol.Phys.* 54, 1335 (1985).
- [6] M.Schlenkrich, K.Nicklas, J.Brickmann and P.Bopp, submitted to *Ber.Bunsen.Phys.Chem.* .
- [7] K.Heinzinger and G.Pálinkás, in "Interaction of Water in Ionic and Non-Ionic Hydrates", edited by H.Kleeberg, Springer Verlag Berlin 1987.
- [8] S.H.Lee, J.C.Rasaiah and J.B.Hubbard, *J.Chem.Phys.* 86, 2383 (1987).
- [9] J.G.H.Joosten, in "Thin Liquid Films" edited by I.B.Ivanov, 569, Marcel Dekker Inc. 1988.
- [10] E.Egberts and H.J.C.Berendsen, *J.Chem.Phys.* 89, 3718 (1988).
- [11] G.Jancsó, P.Bopp and K.Heinzinger, *Chem.Phys.* 85, 377 (1984).

CONFORMATIONAL STUDY OF THE POLAR HEAD OF THE AMPHOTERICIN B IN THE ISOLATED STATE AND IN THE PRESENCE OF WATER MOLECULES.

J. BERGES¹, J. CAILLET¹, J. LANGLET¹, N. GRESIL², M. HERVE³,
 C. GARY-BOBO³.

¹Dynamique des Interactions Moléculaires, Université Pierre et Marie Curie, 4 Place
 Jussieu, 75005 Paris (France)

²Laboratoire de Biochimie Théorique associé au CNRS, 13 Rue Pierre et Marie Curie,
 75005 Paris (France)

³Service de Biophysique, Département de Biologie, CEN Saclay, 91191 Gif-sur-Yvette
 (France)

SUMMARY.

The purpose of this study is to determine of the conformation of the zwitterionic polar
 head of Amphotericin B in both isolated and hydrated states.

INTRODUCTION

There has been, in recent years, a burgeoning interest concerning the biochemical
 mode of action of polyene macrolide antibiotics which appear to be promising as
 antifungal agents. One member of this group, Amphotericin B (AMB, fig. 1), is most
 commonly used in therapy (ref. 1).

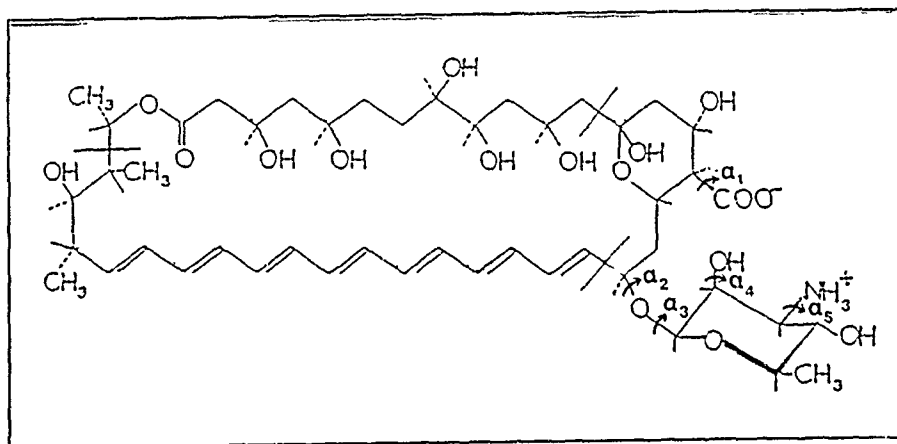


Figure 1. Amphotericin B
 Definition of the fragments and the degrees of rotational freedom

These antibiotics reversibly induce permeability only to monovalent cations in both micro-organisms and animal cells (ref. 2,3). The generally accepted mechanism of ionophoric action of these molecules is based on the assumption that they form complexes with sterols, thus there is a competition between these molecules and phospholipids in the membrane (ref. 4).

It was shown that the polyene macrolides entering the membrane align themselves in a position parallel to sterols and phospholipids molecules (ref. 5) with their polar head at the membrane/water interface. In this position, the polar head functional groups (COO^- and NH_3^+) may interact by H-bonding with $\beta\text{-OH}$ of the sterol. Hydrophobic interactions between the steroid nucleus (and its alkyl tail) and the rigid polyenic part of the macrocyclic ring take place. Both interactions contribute to the stability of the sterol-polyene complex (ref. 6).

Recently, the permeability induced by AMB derivatives in large unilamellar lipidic vesicles containing various sterols has been studied using proton-cation exchange method and ^{31}P -NMR spectroscopy by M.Hervé et al (ref. 7). From the results thus obtained, two groups of polyene antibiotics have been distinguished according to their ionophoric properties. It has been noticed that polyenes of group I have a free ionizable carboxyl group but not those of group II. M.Hervé et al.(ref. 7) have formulated an hypothesis on the role of polyene-sterol interactions in the mode of action of polyene antibiotics following the group to which they belong (see fig. 2a,b).

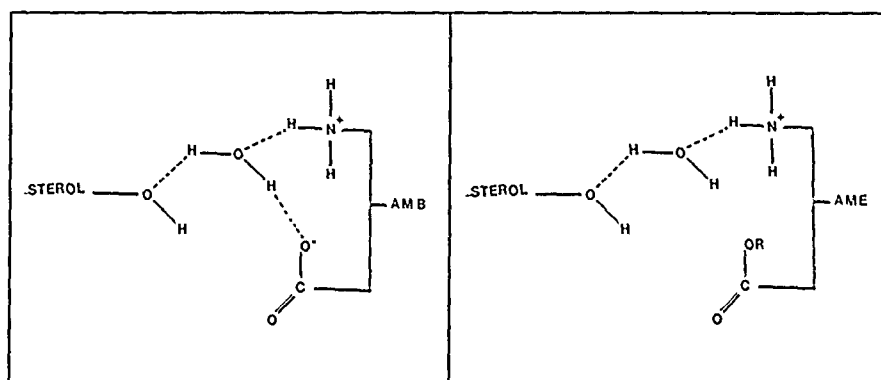


Figure 2. Model of polyene-sterol complex (ref. 7).
a) for group I components b) for group II components.

Such a model needs to be borne out by the study of energetics of polyene-sterol interactions by computer modelling. It may be easily conceived that, for such a theoretical study, a search for the preferred conformation of AMB and derivatives is of prime necessity, but until now experimental data concerning the structure of these groups of molecules are very scarce, except the crystal structure determination of N-iodoacetyl Amphotericin B (ref. 8).

Thus the goal of this work (which represents only a preliminary step in the study of such a problem) is to perform a series of conformational investigations of the zwitterionic form of Amphotericin B. This state of ionization has been chosen because in biological media, the polar head of Amphotericin B is in contact with water. As such, our study will deal with both isolated and hydrated states of Amphotericin B.

1 METHOD

Both intra and intermolecular energies have been calculated simultaneously using the SIBFA method (*Sum of Interactions Between Fragments computed Ab-initio*).

1) Intermolecular energy

The procedure is based on the use of additive components of the intermolecular interaction energy (ref. 10) the expressions of which are fitted in such a way to satisfactorily reproduce the results of *ab initio* SCF supermolecule calculations on small complexes:

$$E_{inter} = \sum_{k=1}^N \sum_{j=i+1}^N E_{inter}(i,j) \quad (1)$$

where N is the total number of molecules in interaction, with

$$E_{inter} = E_{MTP} + E_{POL} + E_{REP} + E_{DISP} + E_{CT} \quad (2)$$

- E_{MTP} : denotes the electrostatic interaction energy between the multipolar expansions (ref. 11) of the *ab-initio* electron density (calculated within an adapted minimal basis set (ref. 12) of the interacting molecules.

- E_{POL} : is the corresponding polarization component.

- E_{REP} and E_{DISP} : are the repulsion and dispersion contributions respectively.

- E_{CT} : is the charge transfer contribution.

2) Intramolecular energy

In the SIBFA method, a large molecule is built out of constitutive molecular fragments separated by single bonds (ref. 13). In fact one calculates the variation of the conformational energy as a sum of inter- fragments interaction energies:

$$\Delta E_{intra} = \sum_{i=1}^N \sum_{j=i+1}^N E_{inter}(i,j) \quad (3)$$

where N is now the number of fragments.

E_{inter} is calculated as a sum of the four first contributions given in eq.(2), plus a term denoted E_{tor} , which is a transferable torsional energy contribution, calibrated for elementary rotations around single bonds (for more details concerning this method see ref. 10,13).

As it was emphasized by Gresh et al (ref. 13), in order to calculate the charge distribution of the whole molecular entity, the constitutive fragments could be defined following two different procedures:

- a) One may compute the multipoles for the fragments which are used in the calculation of inter-fragment interactions.
- b) One may compute multipoles of the largest fragments that can be calculated within an ab-initio framework and then define subfragments for calculating conformational energy changes. The choice between these two procedures is critical and depends on the molecules and fragments studied. This point will be discussed in the Appendix.

As an evaluation of the solvent effect, we have only taken into consideration 'Hydration water' molecules, i.e. the ones which are very close to the solute and thus interact very strongly with it. In order to estimate the 'Hydration energy' (ΔE_{Hydra}), it may be supposed that each water-solute interaction (E_{w-s}) replaces a water-water interaction (E_{w-w}), so that:

$$\Delta E_{Hydra} = \sum_{N_w} E_{w-s} - N_w E_{w-w}$$

N_w being the number of 'Hydration water' molecules. We have used the value of E_{w-w} calculated with the SIBFA method. We are aware that ΔE_{Hydra} only represents part of the total solvation energy in water, but such a study should give an eventual insight into possible intramolecular conformational changes due to these strong water-solute interactions.

II CONFORMATION OF THE POLAR HEAD OF AMB

A - Isolated state

In this present study, all bond lengths and bond angles were fixed to the values obtained from X-ray study (ref. 8). Because of the presence of a conjugated double bond system, we consider that the heptaenic macrolactone ring remains rigid and therefore independent of the surrounding medium; this has been confirmed by Perun and Egan (ref. 9), in the case of erythromycin, another macrolide antibiotic, the conformation of macrolactone ring in solution does not change when compared with crystal state. Hence we have kept this part of the molecule within the conformation established in the crystal and we have been interested only by the flexible polar head. Furthermore we assume that the heptaenic macrolactone ring of Amphotericin B does not influence the conformation of the polar head part of the molecule. In effect preliminary calculations performed when taking (or not) into account this 'rigid tail' have led to strictly identical results.

The conformational energy is expressed as a function of the five variable dihedral angles α_i ($i=1,5$) defined in fig. 1, choosing the geometrical arrangement determined in the crystal as an initial guess (conformation A):

a) We have calculated different conformational energy submaps $E = f(\alpha_2, \alpha_3)$ by means of systematic variations of α_1 dihedral angle values. We have first kept α_4 and α_5 fixed respectively at 0.0° and 150.0° . The deepest energy minimum thus obtained (conformation B₂) has been refined by an automatic minimization process involving the five variable dihedral angles simultaneously. This process has been repeated for different values of α_4 and α_5 .

b) We have also performed direct minimization processes involving simultaneously the five torsional angles.

Our conformational investigation using either strategy led to nearly identical results from both a geometrical and energetical point of view.

The two minima (denoted C₁ and C₂) we have obtained are very close in geometry, and mainly differ by the value of α_4 dihedral angle defining the position of the sugar hydroxyl hydrogen (see Table I).

Conformations	α_1	α_2	α_3	α_4	α_5
A	67.7	272.4	142.1	0.0	150.0
B ₁	67.7	302.4	117.1	0.0	150.0
B ₂	187.7	292.4	92.1	0.0	150.0
C ₁	183.8	292.6	92.0	315.4	175.2
C ₂	202.7	285.7	83.5	220.0	177.3

Table I. Values of dihedral angles defining different Amphotericin conformations in isolated state.

Conformation A determined in the crystal is computed to be 34.4 Kcal/mole above our minimum C₁ and mainly differs by the values of α_2 and α_3 angles.

Analysis of the submap $E = f(\alpha_2, \alpha_3)$ calculated with α_1 fixed at its value in the X-ray structure showed that only a restricted area is allowed in the $\alpha_2 - \alpha_3$ conformational subspace. It resulted in a *unique* minimum, with the 24.4 Kcal/mole energy stabilization proceeding from a simultaneous change in α_2 and α_3 angles (conformation B₁ in Tables I,2). Varying the value of α_1 angle does not change drastically the shape of our submaps. We observed that the location of the *unique* minimum is slightly shifted in the $\alpha_2 - \alpha_3$ subspace (conformation B₂) with an additional stabilization energy of 6 Kcal/mole.

Confor- mations	ΔE_{Repu}	ΔE_{Disp}	ΔE_{El}	ΔE_{Pol}	ΔE_{Tot}
B ₁	18.1	-11.3	-25.9	0.1	-24.4
B ₂	21.8	-18.7	-33.8	-0.7	-30.3
C ₁	25.1	-21.3	-36.6	-0.8	-34.4
C ₂	23.7	-22.3	-27.6	-1.3	-28.0

Table 2. Variation of intramolecular energy ΔE_{Tot} (and its components).
 Energy of conformation A is taken as the zeroth energy.
 Subscripts: *Repu*, *Disp*, *El*, *Pol* for Repulsion, Dispersion, Electrostatic
 and Polarisation respectively. ' 1 values in Kcal/mole).

The conformation of the polar head in the zwitterionic form is mainly governed by electrostatic forces (see Table 2). The stabilization of the closed conformations (C₁ or C₂) is due to an array of three *intramolecular H-bonds* (see fig.3a-b), one of them connecting the lactone ring and the sugar moiety.

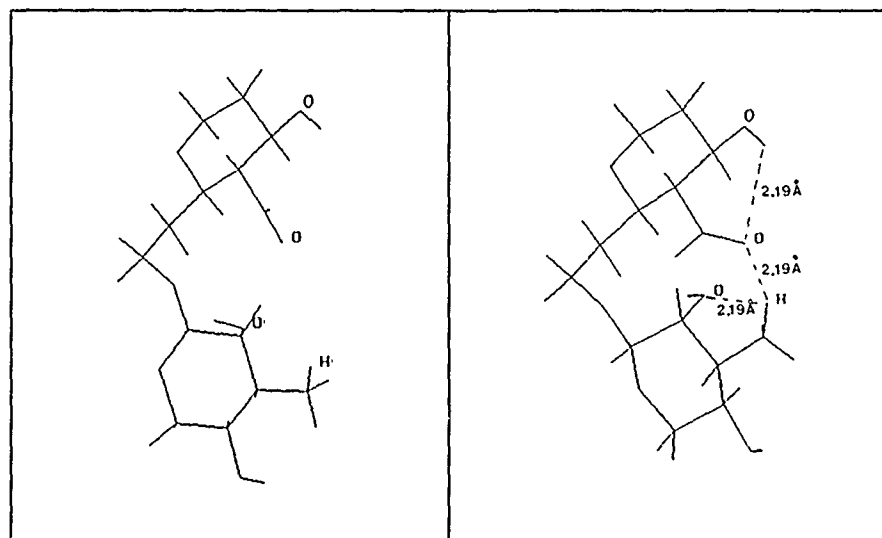


Figure 3. Amphotericin conformations in the isolated state:
 a) from RX data (A). b) Calculated (C₁)

The preferential stabilization of the 'folded' conformation with respect to A was confirmed by SCF-Ab initio calculations with our minimal 'adapted' basis set (ref. 12)

performed on both A and B₂ geometries: B₂ has been calculated more stable than A by 22 Kcal/mole. A more quantitative agreement was not anticipated since it is well known that SCF method does not take into account correlation effects; nevertheless we think that the qualitative agreement is very satisfactory.

B In water

- In a first step, five water 'hydration' molecules were taken into account. Minimization of the sum of inter and intramolecular energies was performed on the crystal conformation (A) and the two folded conformations (C₁ and C₂). The presence of these water molecules does not influence *significantly* the geometrical arrangement of both C₁ and C₂ structures, the intramolecular energy of which remains almost unchanged. Optimizing A conformation leads to a small change in the α_3 dihedral angle value (125° instead of 142°) leading to a structure denoted A' which is intermediate between the 'open' A and the 'folded' C ones, similar to the B₁ conformation (table 1). From an intramolecular point of view, we note a 10.0 Kcal/mole stabilization of A' with respect to A conformation; thus, when taking into account the intermolecular energy, the *total* energy difference between A' and C₁ hydrated molecules favouring C₁ is reduced to 5.8 Kcal/mole (Table 3).

Compared Conformations	N _w	ΔE_{Intra}	ΔE_{Inter}	ΔE_{Tot}
C ₁ / A'	5	-13.6	7.8	-5.8
C ₁ '' / A''	9	-10.8	9.4	-1.4
C ₂ '' / A''	9	-10.7	12.1	1.4

Table 3. Variations of total energy ΔE_{Tot} of hydrated structures (and its components). N_w is the number of water molecules
Subscripts: *Inter*, *Intra* for Intramolecular and Intermolecular.
 ΔE_{Inter} represents the energy difference between ΔE_{Hydra} (defined in Section I) of related conformations. (All values in Kcal/mole).

- In a second step, four additional water molecules were added in order to saturate the polar head. After both intra and intermolecular optimizations, we noted an additional 6 Kcal intramolecular energy gain for A'' conformation. Overall, when considering the *total* energy of the system including the nine water molecules, comparable stabilities (to within 2 Kcal/mole) are derived for A'', C₁'', and C₂'' hydrated structures (Tables 3 and 4). (In Table 3, ΔE represents the energy differences between the C' and A' (respectively C'' and A'') conformations, energy of A' (respectively A'') is taken as the zeroth energy).

Confor- mations	α_1	α_2	α_3	α_4	α_5
A"	79.4	284.8	125.5	7.4	162.4
C'' ₁	188.8	288.9	89.3	36.9	151.4
C'' ₂	213.4	283.8	87.1	215.2	180.1

Table 4. Values of dihedral angles defining different Amphotericin conformations surrounded by nine water molecules.

Fig 4a-b illustrate the position of the nine water molecules surrounding the polar head of A" and C''₂ conformations.

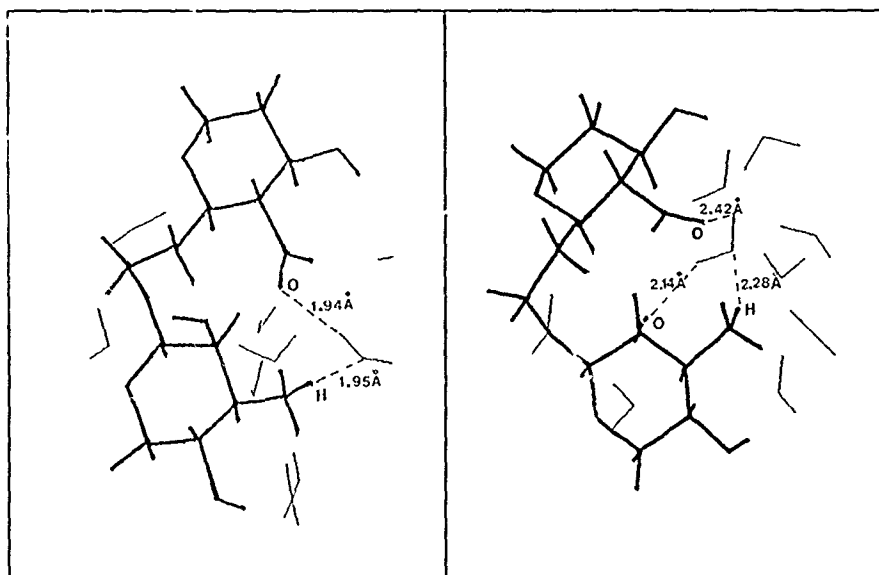


Figure 4. Polar Head of AMB with 'Hydration water' molecules.

a) Conformation A" b) Conformation C''₂

AMB is represented by heavy line. Some molecular distances between AMB and water are represented by dashed lines.

It may be noticed that the hydrated structure still maintains the intramolecular H-bonding which occurs in water-free zwitterion.

As a striking result, the occurrence of a H-bonded bridged water configuration involving : a) COO⁻, NH₃⁺ and OH groups in C''₁ and C''₂ ; b) COO⁻ and NH₃⁺ in A" is noteworthy.

At this stage, we wish to emphasize that our results are distinct from the one obtained by Rinnert and Maigret (ref. 14) who found that the conformational behaviour of

Amphotericin B is independent of the electrical state of the molecule itself and of its surrounding medium. This has led us to investigate whether such a result may not stem from the electrostatic energy contribution, which was computed in ref. 14 by means of Pariser-Parr-Del Re charges. For that purpose, we recomputed the conformational energy differences between A and C conformations now using Mulliken atomic charges rather than the multipolar expansion. These computations (unpublished) again yielded a preferential stabilization of C.

CONCLUSION

The results of our computations show that in the five-dimensional conformational space, the geometrical structure of Amphotericin B is principally governed by two dihedral angles α_1 and α_3 . In the isolated state, the minima corresponding to folded conformations, stabilized by intramolecular H-bonds, remain stable in presence of hydrating water molecules. In the hydrated state, another energetical minimum was found, characterized by: a) a distinct arrangement of water molecules and b) a geometrical structure intermediate between the folded and the open ones. These three structures were found nearly isoenergetic. These results thus exemplify the effects of environmental factors on the conformational preferences of this molecule. In fact, the system manages in order to stabilize its total intra plus inter molecular energy. Thus it is not to be wondered that the geometrical structure we have determined in isolated or hydrated states are different from the one established in the crystal. Let us recall here that the crystal structure determination of Amphotericin B is actually related to a neutral derivative, in which the ammonium is substituted by a bulky N-iodoacetyl substituent. Furthermore, this study enabled us to characterize a particular binding arrangement for one water molecule, bridging together the COO^- and NH_3^+ groups (and eventually sugar OH group). This disposition is reminiscent of the representation recently proposed by Hervé et al. (ref. 7). A complementary (or competitive) H-bond may occur between this water molecule and the β OH group of a sterol molecule in complexes of Amphotericin B with sterols, and work is under progress along these basis.

APPENDIX

We will show that the choice of the constitutive molecular fragments used in the calculation of the charge distribution of the whole molecule is not without consequence upon the final quantitative results. We have derived:

- a) multipoles of the fragments used in the calculation of inter-fragment interaction: M_a .
- b) multipoles of the polar head within two conformations: the folded one B_2 and the open one A.

These two different multipolar distributions are denoted as Mb2 and Mb1 respectively.
 - We have calculated the intramolecular energy difference between both the folded (B_2) and open (A) conformations using these three multipolar distributions.

Fragmentations	ΔE_{El}	ΔE_{Intra}
Ma	-38.3	-37.5
Mb1	-39.4	-34.9
Mb2	-33.8	-30.3

Table 5. Variation of intramolecular energy ΔE_{Intra} for the different fragmentations.
 Energy of conformation A is taken as the zeroth energy.
 Subscripts: *El* for Electrostatic. (All values in Kcal/mole).

Our results listed in Table 5 clearly show that Ma, Mb1, Mb2 lead to the same qualitative conclusions namely the greatest stability of the folded conformation B_2 , but the electrostatic component (and thus the intramolecular energy) depends on the choice of the fragments used in the calculation of the multipolar distribution. We can reasonably think that Mb2 multipoles are the most accurate ones, since they have been obtained from a conformation not involving any intramolecular H-bonds; this is not the case of Mb1 distribution which should reflect a quite important charge transfer between COO^- and NH_3^+ (the ones involved into an intramolecular H-bond), therefore Mb1 multipoles are characteristic of the conformation of the fragment and in agreement with Gresh et al. (ref. 13). We think that this would induce a 'bias' in conformational calculations. On the other hand we have verified that the 7.5 Kcal/mole overestimation we have obtained when using the Ma multipolar distribution mainly results from a *overestimation* of the negative charge on COO^- (-0.874 e^- instead of -0.728 e^- calculated with Mb1). It appears that care should be exercised when ionized fragments are involved in a molecule under study since it is well known that the charge spreads beyond the ionized group. As a conclusion we may reasonably think that accurate multipolar distribution should be obtained if they are calculated from a *large* fragment within a *fully extended* conformation.

ACKNOWLEDGEMENTS

This work was supported in part by Convention 109 RO between University and CEA. The authors thank the Groupement Scientifique IBM-CNRS 'Modélisation Moléculaire' for its financial support.

REFERENCES

1. J. Kotler-Brajtburg, G. Medoff, G.S. Kobayaski, S. Boggs, D. Schlessinger, R.C. Pandey and K.L. Rinchart, Classification of polyene antibiotics according to chemical structure and biological effect, *Antimicrob. Agents Chemother.* 15 (1979) 716-722.
2. B. Malewicz, H.M. Jenkin and E. Borowski, The repair of membrane alteration induced in baby hamster kidney cells by polyene macrolides antibiotics, *Antimicrob. Agents Chemother.* 19 (1981) 238-247.
3. B. Malewicz and E. Borowski, Energy dependence and reversibility of membrane alterations induced by polyene macrolide antibiotics in *Chlorella vulgaris*, *Nature* 281 (1979) 80-82.
4. M. Saint-Pierre Chazalet, C. Thomas, M. Dupeyrat and C.M. Gary-Bobo, Amphotericin B-Sterol complex formation and competition with egg phosphatidylcholine: A monolayer study, *Biochim. Biophys. Acta* 944 (1988) 477-486.
5. N. Ockman, Interaction of Amphotericin B with monolayers of egg lecithin and cholesterol: Polarized absorption spectra, *Biochim. Biophys. Acta* 345 (1974) 263-282.
6. E.J. Dufourc, I.C.P. Smith and H.C. Jarrell, Amphotericin B and model membranes. The effect of Amphotericin B on cholesterol containing systems as viewed by ^2H -NMR *Biochim. Biophys. Acta* 776 (1984) 317-329.
7. M. Hervé, J.C. Debouzy, E. Borowski, B. Cybulska and C.M. Gary-Bobo, The role of the carboxynyl and amino groups of polyene macrolides in their interactions with sterols and their selective toxicity. a ^3P -NMR study. *Biochim. Biophys. Acta* 980 (1989) 261-272.
8. P. Ganis, G. Avitabile, W. Mechlinski, C.P. Schaffner, Polyene macrolide antibiotic Amphotericin B. Crystal structure of the N-Iodoacetyl derivative. *J. Amer. Chem. Soc.* 93 (1971) 4560-4564.
9. T.J. Perun and R.S. Egan, The conformation of erythromycin aglycones, *Tetrahedron Lett.* (1969) 387-390.
10. N. Gresh, P. Claverie and A. Pullman, Intermolecular interactions: Reproduction of the results of *Ab initio* supermolecule computations by an additive procedure, *Int. J. Quantum Chem. Symp.* 13 (1979) 243-253.
11. F. Vigné-Maeder and P. Claverie, The exact multicenter multipolar part of a molecular charge distribution and its simplified representations, *J. Chem. Phys.* 88 (1988) 4934-4948.
12. H. Berthod and A. Pullman, Molecular potential, cation binding, and hydration properties of the carboxylate anion. *Ab initio* studies with an extended polarized basis set, *J. Comput. Chem.* 2 (1981) 87-95.
13. N. Gresh, P. Claverie and A. Pullman, Theoretical studies of molecular conformation. Derivation of an additive procedure for the computation of intramolecular interaction energies. Comparison with *Ab initio* SCF computations, *Theoret. Chim. Acta* 66 (1984) 1-20.
14. H. Rinnert and B. Maigret, Conformational analysis of Amphotericin B I, Isolated molecule. *Biochim. Biophys. Acta* 101 (1981) 853-860.

TRANSPORT IN BIOLOGICAL MEMBRANES . MODELISATION AND EXPERIMENTS.

E. GARCIN¹, F. BAROS¹, J.C. ANDRE¹, D. DAVELOOSE², J. VIRET², M.L. VIRIOT¹
and M. DONNER³

¹ GRAPP-DCPR-UA 328 of CNRS, ENSIC, 1 rue Grandville, BP 451, F-54001 NANCY Cedex

² CRSSA, Laboratoire de Biophysique, BP 87, F-38702 LA TRONCHE Cedex

³ U 284 of INSERM, Plateau de Brabois, CO 10, F-54511 VANDOEUVRE LES NANCY Cedex

SUMMARY

Studies of biological membranes and, more in particular, of membrane dynamics, needs different physical techniques of investigation, allowing to define an order parameter or some lateral diffusion coefficients. For simple model membranes (mixtures of pure phospholipids and purified proteins), we have shown that proteins act as obstacles for the transport properties of lipids. Experiments are issued from three complementary techniques: time resolved fluorescence spectroscopy (quenching of fluorescence of pyren), electronic spin resonance (ESR) and differential scanning calorimetry. Modeling takes into account some concepts of liquid state physics in pseudo-two-dimensional systems as well as the classical Smoluchowski equation of transport.

INTRODUCTION

Biological membranes play an important role in the running of life. Their functions allow them to catalyse chemical reactions as well as to take part to transport phenomena. They are mainly made up with phospholipids, proteins and glucids. The interaction between proteins and lipids is a subject of a great interest in view of a better understanding of such mechanisms. As an example, it is known that the activity of proteins is maximum when they are in presence of a given heterogen composition of lipids. Moreover, the proteins-lipids interactions are supposed to influence the lateral diffusion in membranes. But the characterization of the action of each component is quite hard, because of the complexity of the composition of membranes. For this reason, our approach has been to study first the properties of simple membranes model, made up with an unique phospholipid and then to introduce in them purified proteins.

In order to define some microscopic properties of membranes, we used two classical parameters which are the order parameter S (Zwetkoff parameter) and the lateral diffusion coefficient D . They are directly issued from experiments (fluorescence quenching for D , ESR for S) and their variations with the experimental conditions are supposed to have a real physical significance. Of course, the assumption has been made that the probes don't bring any perturbation

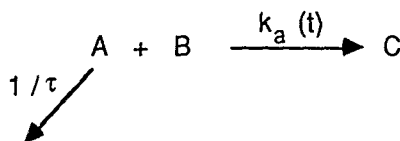
in the membranes and that they can be modelised in the same way as the phospholipids.

The characteristic times explored by each method determine the kind of physical phenomenon seen by them. For fluorescence spectroscopy using pyren as probe, this time is about 10^{-7} s and for ESR about 10^{-8} - 10^{-9} s. This last method will give information about the rotation of small molecules, at the immediate vicinity of the probe. Pyren will explore larger domains and give informations on the lateral diffusion of phospholipids.

1-LATERAL DIFFUSION COEFFICIENT MEASUREMENT

Case of pure phospholipidic membrane

Fluorescence techniques can be mainly classified in two categories : photobleaching and fluorescence quenching. All of them give information on the lateral diffusion coefficient, the last one having the advantage not to modify the membrane geometry during the measurement. The kinetic scheme is as follows:



with τ the natural lifetime of A and $k_a(t)$ the apparent rate constant of the reaction. In this work, we used pyren as B and electronically excited pyren as A.

A model for diffusion limited reaction in two dimensions has already been developed (ref. 1) and we recall here the main features. The basic idea is the adaptation of the classical three-dimensions Smoluchowski's model by considering an homogeneous system of cylindrical molecules B around a given cylindrical molecule A. Solvent is supposed to be a continuum in two dimensions and, if one assumed that each collision between an A and a B molecule leads to reaction, it is possible to calculate the apparent rate constant of the reaction, $k_a(t)$, by solving the classical equation :

$$\frac{\partial \phi}{\partial t} - \left(D \nabla^2 \phi + \frac{D}{k_B T} \nabla \phi \nabla X + \frac{D}{k_B T} \nabla^2 X \right) \quad (1)$$

in which : ϕ is the ratio of the configurational distribution function of B to the mean concentration of B,
 D is the mutual diffusion coefficient of A and B,
 k_B is the Boltzman constant,
 T is the absolute temperature,
 X is the apparent potential, calculated from the radial distribution function into two dimensions, $g(r)$, according to : $X(r) = -k_B T \ln(g(r))$

Of course, some refinements can be made concerning :

- the thickness of the bilayer membrane,
- the curvature of the membrane,
- a random distribution of reactants in the membrane, according to a statistical law of Poisson type.

All of them lead to an apparent rate constant such as :

$$k_a(t) \approx 2\pi N D f \sigma \left(\frac{1}{2} + \frac{2\sigma}{\sqrt{\pi D t}} \right) \quad (2)$$

with f the corrective factor issued of the preceeding refinements, σ the reaction distance of A and B and N the Avogadro number.

The knowledge of the apparent rate constant allows to assume an analytical form for the fluorescence decay of A in the simple following way :

$$\frac{d[A]}{[A]} = - \left(\frac{1}{\tau} + k_a(t) [B] \right) dt \quad (3)$$

leading to :

$$[A(t)] = [A(t=0)] \exp(-\alpha t - \beta \sqrt{t}) \quad (4)$$

in which α and β are linear functions of the concentration $[B]$, depending on the diffusion coefficient and the reaction distance.

Experiments are made using the classical single photon counting technique, described elsewhere (ref. 2). Pyren is incorporated in vesicles and solutions are degased by bubbling argon. Decays are deconvoluted by using a fast Fourier transform algorithm and then fitted by the equation (4). Slopes of α and β versus $[B]$ give D and σ . Figure 1 shows such experimental results for α in different kind of membrane constitutive phospholipids, differing each other by the length of the carbon chain. All experiments are made above the transition temperature, at which membranes are assumed to have properties of a liquid. Under this temperature, coefficients of diffusion are too small and no variation of α versus $[B]$ is observed, as illustrated on figure 1 in the case of DPPC at 25°C. This is of course a limitation of this method. Table 1 summarises the values of D obtained at different temperatures, for different kinds of phospholipids. Two interesting features appear clearly :

- for a given phospholipid (DLPC), the diffusion coefficient increases with the temperature, indicating that membrane acts exactly as a liquid above the transition temperature. This fact was already observed by other authors using different experimental technics (ref.3) and this is in favour of the predictive aspect of our model.

- for a given increase of temperature above the transition temperature (about

10°C), the diffusion coefficient increases with the length of the carbon chain, indicating a greater fluidity in the region where pyren moves. Such a fact seems to be in favour of a localization of the probe quite far from the polar heads of phospholipids.

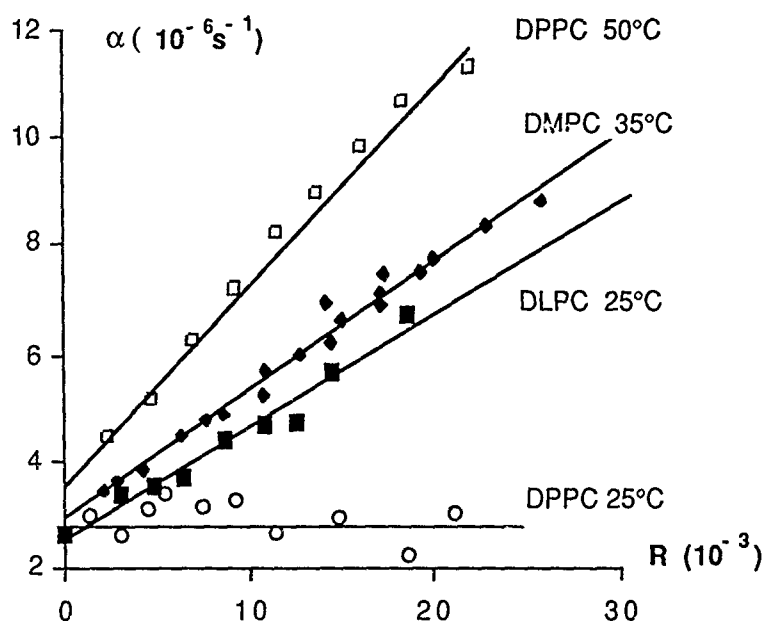


Fig.1. Evolution of α (see text) versus the ratio $R = (\text{pyren})/(\text{phospholipids})$ for different kinds of phospholipids (see nomenclature) at different temperatures.

Phospholipid	number of carbons	T (°C)	D (10^{-7} cm ² /s)
DLPC (0°C)	12	15	1.7
		25	2.9
		35	4.4
		50	5.9
DMPC (24°C)	14	35	3.1
DPPC (41°C)	16	50	6.3
DSPC (59°C)	18	67	6.4

Table 1. experimental values of D (transition temperatures are given into brackets).

Case of mixture phospholipids-proteins

We use now DPPC as phospholipid and GGT as incorporated protein. The surface of the hydrophobic part of this protein has been estimated at about 375 Å², by comparison with other known proteins(ref.4). It is assumed to be a cylinder,

nearly immobile in front of the fast displacements of pyren. All proteins are assumed to be incorporated in membranes and the diffusion coefficients are measured according to the method described above, for different concentrations of proteins (figure 2). The effect is clearly a reduction of the diffusion coefficient when the surface occupied by proteins increases, indicating that this protein has an hardening effect on the membrane, similar to other peptides or enzyme (ref.5).

This effect has been modelised always on the base of the Smoluchowski equation, by assuming a spatial variation of the diffusion coefficient, caused by the presence of the immobile protein. This model has been described in (ref.6). The main result of it is the prediction of a decrease of the mean diffusion coefficient linearly with the surface occupied by the protein. This straight line is also shown on figure 2, and it appears clearly a satisfying agreement with experiments according to the simplification made, even if the model seems to overestimate D .

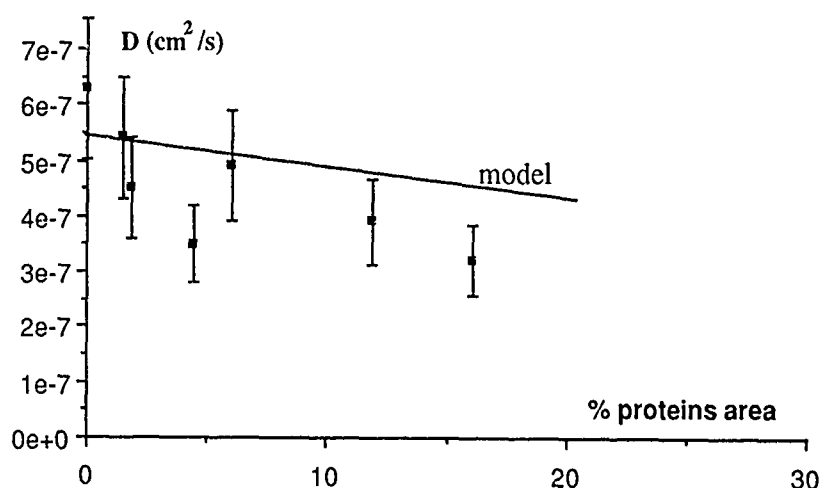


Fig.2. Variations of the diffusion coefficient of pyren with the surface occupied by proteins. Experiments and model.

2-ELECTRON SPIN RESONANCE AND DIFFERENTIAL SCANNING CALORIMETRY EXPERIMENTS

We always study the system DPPC-GGT in which paramagnetic probes 5NS or 16NS are incorporated. These probes are supposed to explore respectively the region near the polar head and near the center of the bilayer. Figures 3 and 4 shows the typical ESR spectrum for each probe, as well as the parameters used to characterize the order around the probe. They also show the variations of these parameters with the temperature, for different concentrations in proteins. The first general remark is that 16NS gives informations under and above the transition temperature, with a clear evidence of the transition. This is not the case for 5NS, for which the transition is very soft, nearly invisible at low concentration in proteins. A second remark concern the disappearance of the pretransition ($T=32^{\circ}\text{C}$)

when proteins are incorporated. Then it appears that the two probes give apparently conflicting results. Indeed, figure 4 shows a fluidizing effect under the transition temperature and a hardening effect above, in agreement with results obtained by fluorescence quenching. On the other hand, 5NS exhibits a hardening effect under the transition temperature.

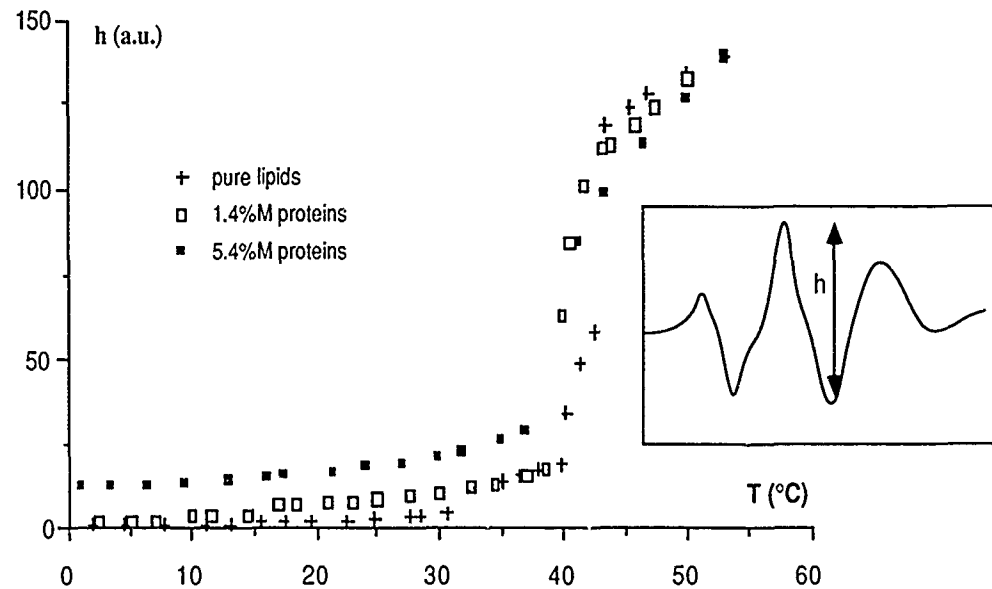


Fig.3. 16NS in membrane: temperature evolution of the parameter h for three concentrations of proteins.

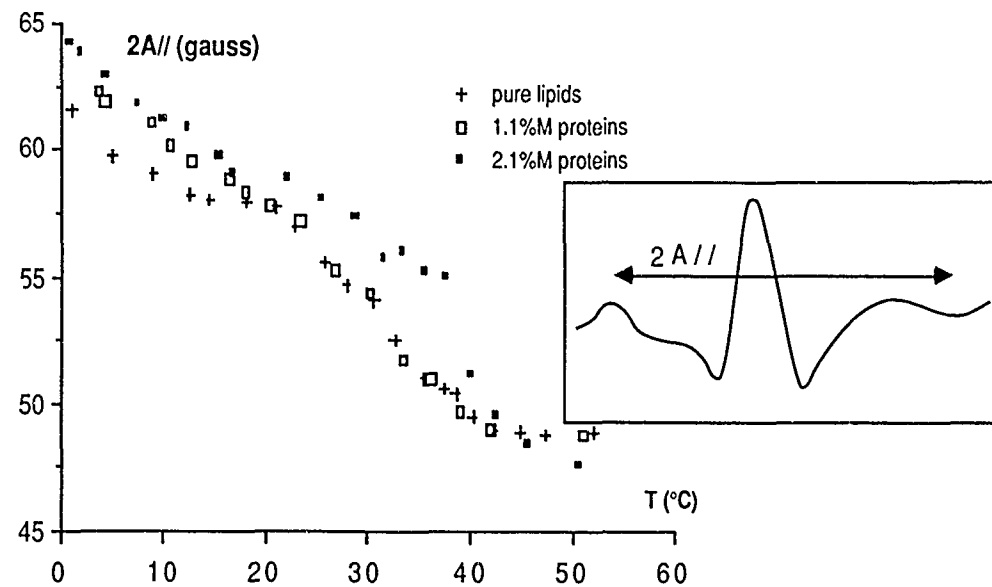


Fig.4. 5NS in membrane: temperature evolution of the parameter $A//$ for three concentrations in proteins.

Moreover, when the concentration of GGT increases, the spectrum of 16NS exhibits an additional peak, witness of the hindrance for some probes to rotate freely.

Complementary experiments have been achieved by using now a macroscopic method, that is differential scanning calorimetry (DSC). It becomes to measure the variations of energy necessary to heat in the same way two samples : buffer as reference and mixture (DPPC + GGT). If a transition is visible, it appears a peak from which one can calculate the enthalpy of transition and a number of cooperativity characterizing this transition. Figure 5 shows the decrease of these two quantities when the concentration in proteins increases.

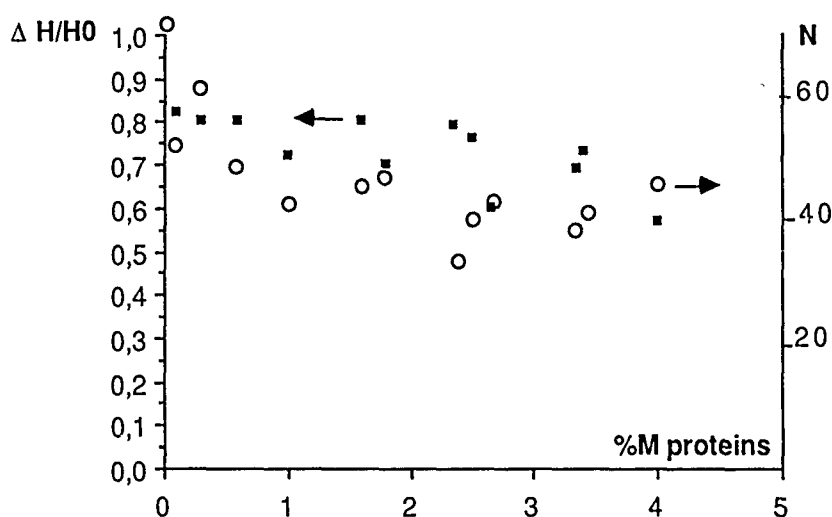


Fig.5. Variations with the concentration in proteins of the relative enthalpy and of the number of cooperativity (N) for multilamellar vesicles of DPPC + GGT.

3-DISCUSSION AND CONCLUSION

Measurement of lateral diffusion coefficient of pyren allow to characterize the lateral displacement of lipids without assuming anything on the membrane fluidity when proteins are incorporated. Indeed, these one can decrease the collision frequency of two lipids by two different way : they induce a hardening effect in the membrane or they act as immobile obstacles and increase the path of two lipids before they collide. It has been noted (ref.7) that, for a given proteic surface, the "hardening" effect is more important for small proteins, which seems in favour of the second assumption.

ESR results indicate that a certain quantity of probes is "locked" by proteins, during a time which can be estimated to about 10^{-7} s. This phenomenon disappears when temperature increases. This is in favour of a model of lipids bound to

proteins. DSC results give an extra information on the number of bound lipids for one proteins: if one extrapolate for $\Delta H = 0$ the results of figure 5, one obtains about 8 lipids in ring around one protein.

The apparent contradiction between results under the transition temperature obtained with 16NS and 5NS can perhaps been explained in the following way : if one imagine that GGT does not penetrate very deep in the membrane, it is easy to see that rotation of 5NS (near the polar head) will be uneasy (hardening effect) when the rotation of 16NS (near the center of the bilayer) will be nearly free.

The decrease of enthalpy observed in DSC experiments can be explained in different ways :

- lipids are bound to proteins,
- lipids are trapped by aggregates of proteins,
- proteins have a global thermodynamic effect on the membrane.

In the first two cases, bound lipids don't participate to the transition and their number increases with the concentration in proteins. But this phenomenon is independent of the direction of the transition. It means that the same results is observed when temperature is decreased from the liquid to the rigid phase. This is again in favour of the first assumption, since aggregates are assumed not to exist in the liquid phase, but some authors (ref.8) have arguments for the third assumption. If DSC experiments don't allow to give a clear response, the indications of ESR experiments remain in favour of the assumption of lipids bound to proteins.

NOMENCLATURE: DLPC: Dilauroylphosphatidylcholine ;
 DMPC: Dimyristoylphosphatidylcholine;
 DPPC : Dipalmitoylphosphatidylcholine; DSPC : Distearoylphosphatidylcholine
 GGT : gamma-glutamyltranspeptidase; 5NS : 5 nitroxyde stearic acid; 16NS : 16 nitroxyde stearic acid

REFERENCES

- 1 F. BAROS, J.C. ANDRE, *React. Kinet. Catal. Lett.*, 36(1) (1988) 1-7
- 2 D.V. O'CONNOR, D. PHILLIPS, *Time-correlated Single Photon Counting*, Academic Press (1984)
- 3 W.L.C. VAZ, R.M. CLEGG, D. HALLMANN, *Biochem.*, 24 (1985) 781-786
- 4 P. CURATOLO, D. SAKURA, D. SMALL, G. SHIPLEY, *Biochem.*, 16 (1977) 2313-2318
- 5 F. JAHNIG, *Proc. Natl. Acad. Sci.*, 76 (1979) 6361-6365
- 6 J.C. ANDRE, M. BOUCHY, M.L. VIRIOT, M. DONNER, *Immunologie et Cancer, Symposium*, Montpellier (FRANCE) (1981)
- 7 J. EISINGER, J. FLORES, W.P. PETERSEN, *Biophys. J.*, 49 (1986) 987-1001
- 8 M.R. MORROW, J.H. DAVIS, F.J. SHARON, M.P. LAMB, *Biochim. Biophys. Acta*, 858 (1986) 13-20

RELATIONSHIP BETWEEN HYDROGEN BONDING PROPERTIES OF SOME
PHOTOSYSTEM II HERBICIDES AND THEIR ACTIVITY ON TRIAZINE-
RESISTANT AND SUSCEPTIBLE CHLOROPLASTS.

S. CREUZET, B. GILQUIN and J.M. DUCRUET

Service de Biophysique, Departement de Biologie, CEA/INRA
Saclay - 91191 Gif-sur-Yvette (France)

SUMMARY

The hydrogen bonding properties of two classes of photosystem II inhibitors (amide/urea and s-triazine derivatives) were studied by the semi-empirical quantum method AM1, using water, methanol and p-chlorophenol as proton donors. These calculations showed a lower hydrogen bonding capacity of the triazine ring nitrogens compared to the carbonyl of amide/urea inhibitors. However, some discrepancies arose from the comparison with IR spectrometry data.

INTRODUCTION

The herbicides belonging to urea/amide, uracil, s-triazine and cyanoacrylate families inhibit the photosynthetic electron flow by competing with a plastoquinone molecule for binding at a common site, on the D1 protein of photosystem II (PS II). The binding of the plastoquinone in the D1 site forms the secondary acceptor Q_B , which reoxidizes the photochemically reduced primary acceptor Q_A^- (ref. 1).

The role of hydrogen bonding in the mechanism of this inhibition, first postulated from QSAR studies (ref. 2), was demonstrated by X-ray crystallography showing the interactions of the triazine terbutryne with the bacterial reaction center (ref. 3,4). Even though hydrogen bonding is likely to contribute only a part of the whole interaction energy, it seems to be essential for the binding of these inhibitors.

However, the PS II centers are somewhat different from bacterial centers (refs. 3,5) and no PS II crystal structure has been obtained so far. Since amide/urea derivatives (diuron-like inhibitors), unlike triazines, are inactive on bacterial photosynthesis, the mechanism of their binding on the PS II site remains unknown. A homology between the -NH-CO-

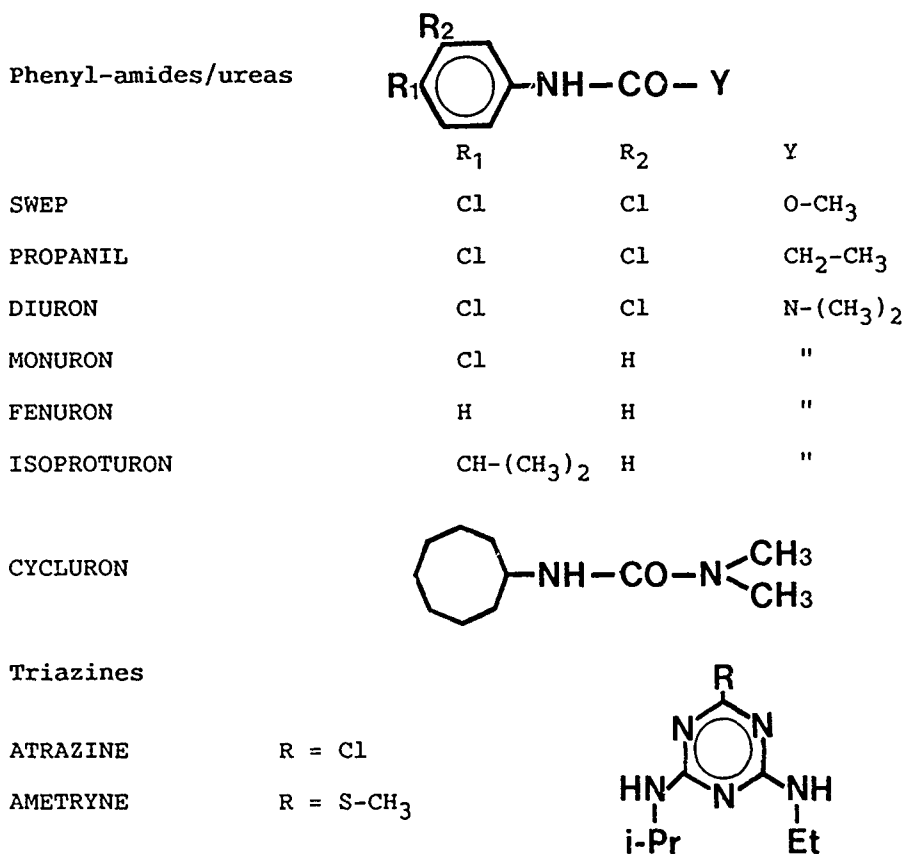


Fig. 1. Structure of the PS II inhibitors studied.

common motif of diuron-like inhibitors and the -NH-CN= pattern present in triazines was proposed long time ago (refs. 2,6) and the evidence of a double hydrogen bond formed by the latter in bacterial reaction centers (ref. 3) suggests a similar mechanism of interaction for the former. However, this implies that the energetically unfavoured *cis* form of the amide would interact with the site. Furthermore, the inhibitory power of diuron-like compounds remains almost the same in triazine-resistant mutants of higher plants, in which the substitution of serine 264 by a glycine in the binding pocket induces a 500 fold decrease in the activity of atrazine (ref. 7).

Since a triazine ring nitrogen or the carbonyl of amide/urea derivatives are supposed to form the interaction

with the site, we have here compared the hydrogen bonding capacities of several structures from these groups of inhibitors with model proton donors such as water, methanol and p-chlorophenol. Calculations were performed *in vacuo* using the semi-empirical quantum chemistry method AM1 and the results were compared with hydrogen bonding measurements by IR spectroscopy in CCl_4 . The size of the systems studied, ranging from 25 to 50 atoms, did not allow reasonable *ab initio* calculations to be performed. Since AM1 seems to give more reliable results than other semi-empirical methods for the molecular properties of the compounds here considered (ref. 8) and for hydrogen bonding calculations in general (ref. 9), it was chosen for this work.

METHODS

Calculations

Structures were constructed and minimized using a SYBYL package including the molecular mechanics TRIPOS 5.1 force field. The resulting sets of coordinates were fully optimized by the Quantum Chemistry Method AM1 1.0 (ref. 10). Atomic charge densities were calculated for the optimized structure.

Infra-Red spectroscopy

Hydrogen bonding between some of the inhibitors studied and p-chlorophenol was measured by IR spectrometry, from the absorbance of the longitudinal vibration of the unbounded phenolic hydroxyl. The solvent was carbon tetrachloride, in cuvettes of 1 cm optical path. Concentrations of phenol and inhibitors were sufficiently low (1 to 4 mM) to avoid self-associations.

RESULTS

Pyridine - H_2O

In their evaluation of AM1 for hydrogen bonding studies, Buemi et al (ref. 10) conclude that this method give better results than other SEQCM methods, although energies are still underevaluated and H bond distances overevaluated, compared to experimental data. We did reproduce these results in the case of the pyridine + H_2O complex by a single full optimization.

Two other procedures were tested in order to determine both H-bond energy and distance (Fig. 2):

- In the first procedure, the geometry was varied by decreasing the N...H...O distance without optimizing either the internal coordinates or the water orientation relative to the pyridine.

- In the second procedure, the whole complex was fully optimized at each step, allowing rearrangement of the water molecule orientation.

The second procedure gave a lower heat of formation at the minimum and a lower hydrogen bond distance. The water molecule orientated such that both protons interacted with the pyridinic nitrogen.

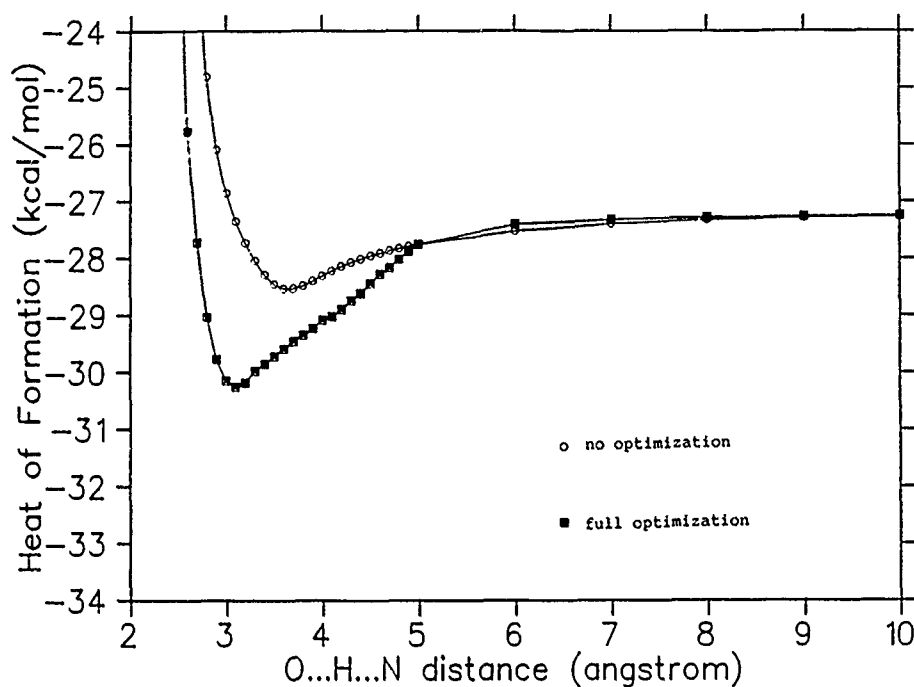


Fig. 2. Variation of the heat of formation of water-pyridine complex during formation of hydrogen bond modelled by different procedures.

Inhibitors - H₂O

These procedures were applied to triazine and amide/urea structures, giving rise to particular features due to the

orientation of the water molecule which may lead to the formation of more than one hydrogen bond.

In the purpose of evaluating the capacity of the inhibitors to form single hydrogen bonds with the proteic site, we fixed the orientation of the water molecule such that only

TABLE 1a.

Model of hydrogen bond formation with water.

MOLECULE	H_f°	δ^-_X	d X..H..O
SWEP	3.986	-0.3806	3.10
PROPANIL	4.022	-0.3406	3.15
DIURON	4.207	-0.3814	3.09
MONURON	4.216	-0.3825	3.09
FENURON	4.251	-0.3845	3.09
ISOPROTURON	4.401	-0.3841	3.08
CYCLURON	4.419	-0.4020	3.07
AMETRYNE	2.373	-0.2718	3.48
ATRAZINE-2	1.513	-0.2864	3.63
ATRAZINE-3	1.969	-0.2822	3.55

TABLE 1b.

Model of hydrogen bond formation with para-chloro phenol.

MOLECULE	H_f°	δ^-_X	d X..H..O
SWEP	4.547	-0.3806	3.08
PROPANIL	4.850	-0.3406	3.08
DIURON	4.978	-0.3814	3.04
MONURON	5.157	-0.3825	3.03
FENURON	5.280	-0.3845	3.05
ISOPROTURON	5.101	-0.3841	3.04
CYCLURON	5.563	-0.4020	3.04
AMETRYNE	2.517	-0.2718	3.40
ATRAZINE-2	2.602	-0.2864	3.48
ATRAZINE-3	2.531	-0.2822	3.69

H_f° (kcal/mol) is the difference between the sum of heat of formation of both molecules and the one of the optimized complexe, δ^-_X is the partial charge on the atom acceptor in the studied molecule. d X..H..O (angstrom) is the characteristic distance in the optimized complexe. Geometry of Hbond is fixed as well as X.H.O angle (180°).

one bond could be formed. Internal coordinates of both molecules and distances between them were then optimized. The results in Table 1 show a clear distinction between $N\cdots H-O$ (atrazine-like) and $O\cdots H-O$ (diuron-like) and a weaker bonding capacity of amides (propanil, swep) compared to ureas, in which the dimethylamino substituent enhances the nucleophilicity of the carbonyl. However, the substituents on the N-phenyl had no influence on the energy of bonds formed by the carbonyl.

Similar results were obtained with methanol instead of water, but the hydrogen bonding energies were too weak to allow a distinction between urea/amide derivatives to be made.

Inhibitors - p-chlorophenol

With parachlorophenol instead of water, as the hydrogen donor, there was still a clearly lower energy for atrazine-like than for diuron-like inhibitors. Variations of bonding energies between amide/urea derivatives were in a wider range than those

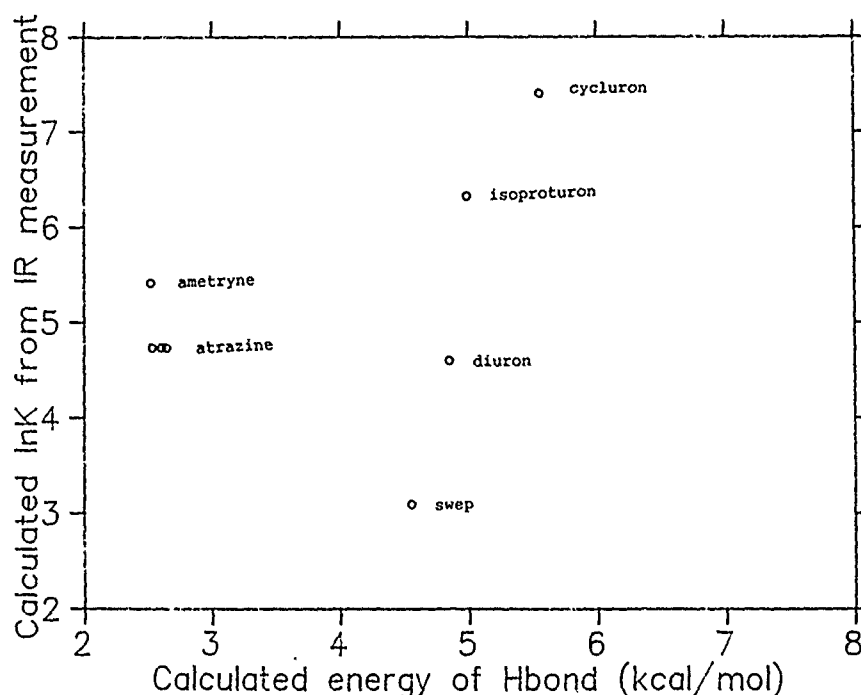


Fig. 3. Relation between the association constant K of p-chlorophenol/inhibitors complexes and the Hbond energy calculated by AM1.

Observed with water and the influence of substituents on the N-phenyl became significant.

IR spectroscopy

In the presence of inhibitors, the intensity of the free OH vibration of p-chlorophenol (3610 cm^{-1}) decreased and an hydrogen-bonded band appeared at lower frequencies only for amides/ureas. For triazines, a featureless absorption was detected, which can be ascribed both to the presence of three unequivalent ring nitrogens and to greater delocalization of the proton (ref. 11).

The IR association constants (K) of atrazine-like and diuron-like compounds with p-chlorophenol fall in the same range. For these latter compounds, the strength of hydrogen bonding, consistently measured by both K values and the frequency shifts of the OH bonded bands (data not shown) indicate a much greater influence of the substituents on the N-phenyl than that predicted by computation (Fig. 3).

DISCUSSION

Hydrogen bonding calculations presented here show a clearly lower proton accepting capacities of the nitrogens of triazine ring as compared to the oxygen of amide/urea and other carbonyl-containing inhibitors.

Some discrepancies arise in the comparison of these results with experimental hydrogen bonding measurements:

- the clear distinction between atrazine-like and diuron-like inhibitors resulting from AM1 calculations is not supported by the IR-measured association constants. However, the shape of the hydrogen bonded OH bands are quite different for these two classes of inhibitors, ascribable to different properties of the $\text{N}\cdots\text{H}-\text{O}$ and $\text{O}\cdots\text{H}-\text{O}$ bonds (ref. 11).

- The influence of the substituents of the N-phenyl on the strengths of hydrogen bonds formed by diuron-like inhibitors was greater when measured by IR spectrometry than when calculated. QSAR studies have already shown that these substituents also influence the inhibitory power through an electronic inductive effect (2).

Differences in the proton accepting capacities of atrazine-like and diuron-like PS II inhibitors could contribute to explain their contrasting behaviour towards triazine-resistance, in the hypothesis of a common mechanism of interaction with the D1 site of inhibition.

REFERENCES

- 1 B.R. Velthuys, Electron-dependent competition between plastoquinone and inhibitors for binding to photosystem II, FEBS Letters, 94 (1981) 277.
- 2 C. Hansch and E.W. Deutsch, Structure activity relationship in amides inhibiting photosynthesis, Biochim. Biophys. Acta, 112 (1969) 381-391.
- 3 H. Michel and J. Deisenhofer, Relevance of the photosynthetic reaction center from purple bacteria to the structure of photosystem II, Biochemistry, 27 (1988) 1-7.
- 4 I. Sinning, H. Michel, P. Mathis and A.W. Rutherford, Characterisation of four herbicide-resistant mutants of *Rhodospseudomonas viridis* by genetic analysis, electron paramagnetic resonance, and optical spectroscopy, Biochemistry, 28 (1989) 5544-5553.
- 5 A. Trebst, The three-dimensional structure of the herbicide binding niche on the reaction center polypeptides of photosystem II, Z. Naturf., 42c (1987) 742-750.
- 6 A. Trebst and Harth, Herbicidal N-alkylated ureas and ring closed N-acylamides as inhibitors of photosystem II, Z. Naturf., 29c (1974) 232.
- 7 K. Pfister and C.J. Arntzen, The mode of action of photosystem II-specific inhibitors in herbicide-resistant weed biotypes, Z. Naturf., 34c (1979) 996-1009.
- 8 S. Creuzet, B. Gilquin and J.M. Ducruet, Relationship between molecular properties of some photosystem II inhibitors and their activity on triazine-resistant and susceptible chloroplasts, Z. Naturf., 44c (1989) 435-443.
- 9 G. Buemi, F. Zuccarelo and A. Raudino, Hydrogen bonding and rotation barriers: a comparison between MNDO and AM1 results, J. Mol. Struct. (Theochem), 164 (1988) 379-389.
- 10 M.J.S. Dewar, E.G. Zebisch, E.F. Healy and J.J.P. Stewart, AM1: a new general purpose quantum mechanical molecular model, J. Am. Chem. Soc., 107 (1985) 3902-3909.
- 11 P.P. Rastogi and G. Zundel, Proton translocation in hydrogen bonds with large proton polarisability formed between a schiff base and phenol, Biochem. Biophys. Res. Commun., 99 (1981) 804.

3D-PATTERN RECOGNITION OF PHARMACOPHORS. APPLICATION OF THE METHOD TO THE COMPARISON MORPHINE-ENKEPHALINE.

C. RICHE¹ and L. MOY².

¹ Institut de Chimie des Substances Naturelles, C.N.R.S., 91190 Gif/Yvette.

² Département scientifique, Aquitaine Systèmes, 92078 Paris la Défense 6.

SUMMARY

A method of 3D pharmacophor recognition has been developed in order to compare 2 molecules of identical biological activity. The PPSP3 program takes into account the receptor topology and proposes hypothesis of pharmacophors based on pure geometrical criteria. A first attempt has been made a few years ago by C. Leroy (ref.1). PPSP3 has been improved in order to consider more complex pharmacophors and fast energetic calculations have been incorporated. Application of the method has been applied to the classical morphine-enkephaline comparison. One of the PPSP3 solutions indicates a pharmacophor of the tyrosine differing from the classical model.

INTRODUCTION

Recent works on opiate compounds have shown 3 types of different receptors : μ (for the morphine), δ (enkephaline), κ (benzomorphan)(ref.2). Since the discovery of enkephalines (ref.3), numerous works have been proposing models of conformation by comparing enkephalines and morphinic compounds. As very small changes in conformation give rise to a high variation of selectivity towards the receptor, most of the people accept the fact that enkephalines, which are typical δ ligands, could interact with the μ site and induce biological effects specific of the μ receptor (ref.4). Starting from the hypothesis of a common interaction at the μ site, PPSP3 program has been applied to the 3D pharmacophor search between morphine (rigid active analog) and Leu-enkephaline (flexible analog). Results show that two molecules may have a common pharmacophor without geometrical backbone superposition. Conformational study has been made on pharmacophors hypothesis using Scheraga program ECEPP (Empirical Conformational Energy Program for Peptides)(ref.5). The most favoured solution of the program has been compared and superimposed with cristallographic structures.

METHODS

Input PPSP3 Datas.

Coordinates of the atoms come from data banks such as Cambridge for small molecules or Brookhaven for peptides. Data are issued from geometrical optimization by molecular mechanics (MM2 program) or MAD (Molecular Advanced Design) when crystalline coordinates of the rigid analog are unknown. In a first step, coding of the molecules have to consider interactions towards the receptor. One molecule is defined by its interaction centers with the hypothetic receptor : classical atoms (oxygen, nitrogen...) play a fundamental role in the interaction and can be defined as active centers. Lone pair of an heteroatom, direction of an aromatic group define a second class of points on the hypothetic receptor and are named "complementary active centers". Choice of the coded atoms is made by the user of the program. Example of the second class of coded points is made with the illustration of two active centers representing a phenyl (Fig.1).

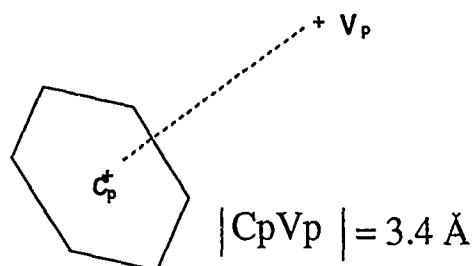


Fig. 1. Illustration of two active centers, phenyl center C_p and its direction V_p representing a phenyl ring (distance between the two points is 3.4 Å).

Pharmacophor Search.a- Basis of the method.

Each study includes one rigid molecule and an active flexible analog (maximum of 6 degrees of freedom). The method is based on the comparison of two samples of active centers, one being fixed, the other one (belonging to the flexible analog) being reevaluated for each new investigated conformation. As each new conformation of the flexible molecule defines a set of active centers, the goal of PPSP3 is to find the common set of points between the two molecules. This superposition (maximum of superimposed active centers) is then kept as a solution of the program. The first step of this comparison is the elimination of the origin of coordinates. The autoconvolution product is the conversion of two active centers A and B (origin O) into a vector AB represented by a norm and a direction. Example of this mathematic procedure applied to three points A, B, C is shown on Fig. 2.

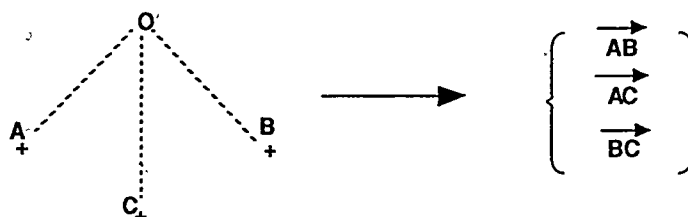


Fig. 2. Autoconvolution product on A, B, C.

In a general way, $N(N-1)/2$ vectors are associated to a set of N active centers (A_1, A_2, \dots, A_N). As the scalar product of AB with AC gives the orientation (angle value α) between the two vectors (Fig. 3), if the correspondance of the X_n values (here $n=3$) between the two sets of active centers is successful, then a pharmacophor of 3 points is found.

$$\left\{ \begin{array}{c} \overrightarrow{AB} \\ \overrightarrow{AC} \end{array} \right\} \xrightarrow{\overrightarrow{AB} \cdot \overrightarrow{AC}} \left[\begin{array}{l} X_1 = AB \cdot AB \\ X_2 = AB \cdot AC \cdot \cos \alpha \\ X_3 = AC \cdot AC \end{array} \right]$$

Fig. 3. X_1 , X_2 and X_3 determine the comparison between 2 sets of vectors, α being the angle value between AB and AC .

This method is easily generalized to a number of n points or active centers in order to recognize complex pharmacophors. The program PPSP3, written in fortran 77, is then able to indicate superpositions of conformations for unlimited number of active centers.

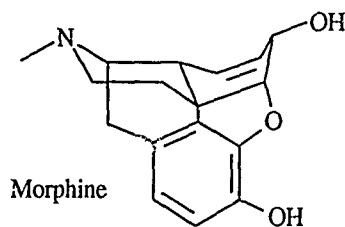
b- Limitation of the solutions.

Reiteration of the precedent method is done for each new conformation of the flexible analog. As small peptides could give rise to a drastic number of conformations, PPSP3 allows the user to run magic numbers technic in order to decrease the number of investigated conformations and proposed pharmacophors. In case of 3 moving dihedral angles with a step of 9° , $(360/9)^3$ e.g. 64000 conformations are required instead of 520 conformations by using the magic numbers. As up to now, conformations were generated in PPSP3 without taking into account energetic values, it is reliable to say that the conformation fitted to the receptor is not far from a local minimum of energy and for this reason, the method cannot ignore the

conformational energy of a molecule. The steric energy is calculated by many molecular mechanics programs such as MM2 from Allinger or other molecular modeling softwares such as MAD (Molecular Advanced Design). Once bond lengths and angles are well established, the Van der Waals interactions (interactions between non-bonded atoms) remain the quickest way for an energetic approach. The Van der Waals energy is then a good criterion for elimination of high energy conformations. The function used for Van der Waals calculation is the well known 6-exponential function of Allinger's programs.

APPLICATION

PPSP3 is applied to a simplified model including the tyrosine of the enkephalines. The dihedral angle values χ_1 and χ_2 will determine the conformation of the tyramine moiety. As the correspondance morphine-tyrosine of enkephalines has appeared to be obvious since the discovery of these pentapeptides, the direct superposition of the two moieties should be one of the PPSP3 solution. The morphine is the basic model of analgesic compounds. It includes a classical pharmacophor represented by the phenylpiperidine group (T shape). The structure is rigid and the crystalline coordinates are well established.



Enkephalines are known to be δ ligands and to induce typical morphinic biological effects by adopting a different conformation when binding the μ receptor. The sequence of the Leu-enkephaline is Tyr-Gly-Gly-Phe-Leu.

a- Coding of the molecules.

The morphine, considered as the rigid molecule is coded with 5 active centers (nitrogen, phenol hydroxyle, center and direction of the phenol). Other carbons of the phenethyl group are also coded. Such a definition will show clearly the superposition of the 7 active centers. By the same way, the tyrosine is coded with the same active centers and includes 3 moving dihedral angles (Fig.4).

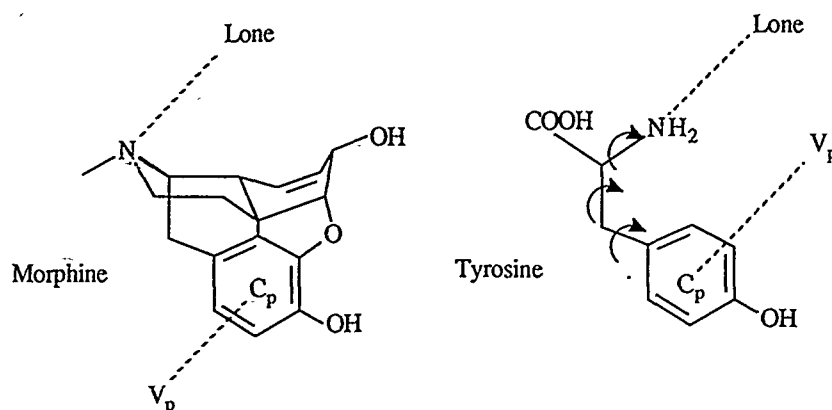


Fig. 4. Coding of the two molecules.

b-PPSP3 pharmacophor search.

The study has been done with a step of 9° for the 3 dihedral angles of the tyrosine and by using the technic of the magic numbers. From this search, 2 solutions corresponding to 2 different pharmacophors appear clearly :

- Pharmacophor 1 or direct superposition : this is the obvious superposition with the correspondance of the 7 active centers (nitrogen lone pairs noted Lone, center and direction of the aromatic rings noted Cp and V_p).
- Pharmacophor 2 : this new conformation shows the superposition of the same active centers. Except the nitrogens and the phenethyl carbons, all the points are superimposed and the directions of the lone pairs point to a common area. A classical superposition similar to this case deals with the QSAR study of the two isomers of 4-phenylpiperidine (Fig.5).

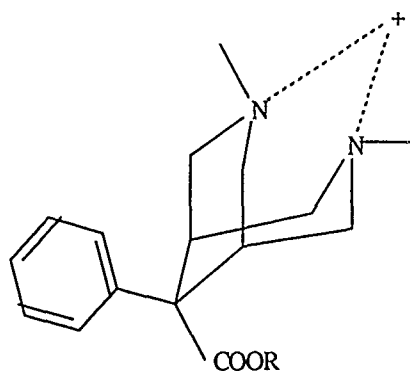


Fig. 5. Superposition of the 2 isomers of 4-phenylpiperidine both inducing analgesia at the morphinic receptor.

In this kind of superposition, the basis of the action of enkephalines towards μ receptor is considerably different from the classical model. In order to check the validity of each solution (pharmacophor 1 or 2), energetic calculations are performed for each conformation.

c- Conformational analysis of solutions by ECEPP.

Conformational analysis of PPSP3 solutions (pharmacophors 1 and 2) is made by empirical calculations of the energy (program ECEPP for Empirical Conformational Energy Program for Peptides)(ref.5). The comparison with a crystalline structure of the Leu-enkephaline will provide more informations to check the previous solutions. The study has been restricted to the first 2 amino-acids of the Leu-enkephaline, e.g. the sequence $\text{NH}_3^+\text{-Tyr-Gly-COO}^-$ and by considering the nitrogen protonated at the pH of the physiological medium. By fixing the tyrosine with a conformation corresponding to pharmacophor 1 or pharmacophor 2, one can evaluate the validity of the model with a minimization of the energy done with conformational constraints.

- Study of pharmacophor 1 (Table 1) : values issued from ECEPP are in kcal. and correspond to the sequence of dihedral angles (Φ , Ψ , χ_1 , χ_2) of the direct superposition. The first serie of angles and energy (23.4 kcal.) comes from the first PPSP3 solution. The constraints imposed by pharmacophor 1 (χ_1 , χ_2 fixed) are then relaxed to obtain a stable conformation (5.5 kcal) next to the initial sequence of pharmacophor 1.

TABLE 1

Energy	Ψ	Φ	χ_1	χ_2
24.3	58	-64	-108	14
5.5	58	-64	-77	104

Energy and dihedral angles for the pharmacophor 1 (first row) are displayed as its corresponding relaxed conformation (second row).

The initial energy (24.3 kcal) is very high compared to the minimized value (5.5 kcal). The final values of χ_1 and χ_2 are quite different from the

initial angle values and the pharmacophor 1 defines a conformation far from the next local minimum.

- Study of pharmacophor 2 : the initial conformation (Φ , -64° , Ψ , -52° , χ_1 , 200° , χ_2 , 84°) has a lower steric energy of 10.1 kcal. The sequence of dihedral angles after relaxation leads to values χ_1 (168°), (χ_2 65°) and a very low energy value of 4.8 kcal. In this case, the energy value is lowest (-14 kcal.) than for pharmacophor 1. Dihedral angles after relaxation are close to the initial values. The pharmacophor 2 defines a more stable conformation than the one corresponding to the first pharmacophor. For pharmacophor 1, the angle χ_2 is close to 0° (14° in the first row of Table 1) ; as the more stable state for the tyrosine in a free rotation is the staggered conformation (relatively to the phenethyl carbons), the pharmacophor 2 reduces the steric interactions. On the other hand, the conformational constraints of the rigid morphine induces a high steric energy value for pharmacophor 1.

Correlation is made with the crystalline structure of Leu-enkephaline for pharmacophor 2.

d- Superposition Leu-enkephaline -morphine.

The pentapeptide Leu-enkephaline trihydrate P212121 has been chosen as it has been crystallized in water (similar to the physiological medium). By comparing the PPSP3 solutions with this stable conformation, the methodology adopted to check our results can be compared to the previous method (comparison with structures optimized by molecular mechanics). The calculation of the dihedral angles Ψ and Φ of the different amino-acids gives the conformation of the peptide after its crystallization (Φ_2 , -60° , Ψ_2 , -30° , Φ_3 , -50° , Ψ_3 , -60° , Φ_4 , -80° , Ψ_4 , 4°). The 3 amino-acids Tyr-Gly-Gly induce an α helice conformation followed by a β turn of type I starting from Phe⁴. The superposition with the pentapeptide is possible by considering the conformation of the second PPSP3 solution (pharmacophor 2) as direct superposition morphine-Leu-enkephaline cannot be obtained because distances nitrogen-phenyl center are too different in the 2 molecules :

- distance N - Cp (tyrosine) = 5.2 Å.
- distance N - Cp (morphine) = 4.5 Å.

For the tyrosine, Φ (-70°), Ψ (125°), χ_1 (178°), χ_2 (47°) are dihedral angles values close to the ones obtained for pharmacophor 2. A consequence

of these results is that Leu-enkephaline and morphine adopt a conformation at the mu receptor different from the classical model. The phenylalanine moiety of enkephaline is necessary for the interaction of the peptide towards the delta site and orientation of this aromatic ring is dominant with respect to the selectivity ; this has been shown by Portoghese with the discovery of new selective delta ligands (ref.6). The hypothesis that the enkephalines could interact with the phenylpiperidine moiety of the morphine has been suggested by Portoghese. As the hybrid of enkephaline, where the tyrosine has been replaced by (-) métazocine (ref.7), is inactive, Portoghese concludes that the tyramine moiety in the opiates does not play the same functional role than for peptides. In fact the inactivity of these compounds could suggest that only the orientation of the phenylpiperidine ring is crucial for the activity at the mu receptor. Such a peptide conformation described for the pharmacophor 2 is then essential for an interaction at the mu site. The conclusion is that tyramine moieties between morphinic compounds and enkephaline peptides could play the same role by adopting different conformations (pharmacophor 1 for the morphine, pharmacophor 2 for Leu-enkephaline).

CONCLUSION

The method of shape recognition used in PPSP3 is useful for the determination of common and complex pharmacophors. The program has been applied to the classical morphine-enkephaline comparison, assuming that the peptide also binds to the mu site. By considering pure geometrical criteria, PPSP3 search has given 2 hypothesis of pharmacophors. The direct superposition of the 2 structures has shown a high steric energy, ECEPP program providing all the energetic informations of the studied conformations. Comparison with the crystalline structure of Leu-enkephaline trihydrate has also been performed. If the first applications of the PPSP3 pharmacophor search have concerned the morphine/enkephaline model, the described method could certainly be applied to another series of drugs.

REFERENCES

- 1 C. Leroy, personal communication, Recherche tridimensionnelle de pharmacophores, Univ. Paris Sud, Orsay, 1983.
- 2 W.R. Martin., C.G. Eades, J. Pharm. Exp. Theor., 1976, 197, 517-532.
- 3 J. Hughes, H.W. Kosterlitz, T. Smith, Life Science, 1975, 16, 1753.
- 4 B.P. Roques, G. Gacel, J.M. Zajac, J. Med. Chem., 1988, 31, 836.
- 5 H.A. Scheraga, Y. Isogai, G. Nemethy, Proc. Natl. Acad. Sci., 1977, 74, 414.
- 6 P.S. Portoghese, K. Ramakrishnan, J. Org. Chem., 1980, 45, 4240.
- 7 P.S. Portoghese, K. Ramakrishnan, J. Med. Chem., 1982, 25, 1425.

MOLECULAR DYNAMICS SIMULATION OF A N-METHYL INHIBITOR BACTERIAL COLLAGENASE

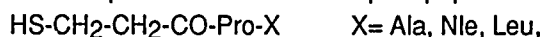
B. GILQUIN ¹, D. PERAHIA ² and V. DIVE ¹

¹ Service de Biochimie, Département de Biologie, et Laboratoire d'Ingénierie des Protéines, CEN-Saclay, 91191 Gif-sur-Yvette Cedex

² Laboratoire d'Enzymologie Physico-Chimique et Moléculaire, Bat 430, Université Paris-Sud, 91405 Orsay Cedex.

INTRODUCTION

Structure activity relationships in a series of mercaptotriptides



which are inhibitors of bacterial collagenase, have shown that the addition of a methyl group onto the amide group of X, enhanced their inhibitory activity (for example when X=Ala the KI value increases from $60 \cdot 10^{-8}$ M to $17 \cdot 10^{-8}$ M). One effect of the methyl group is to restrict the conformational space of the inhibitor. In order to get a deeper insight into the conformational perturbation due to the N-methylation, the conformational properties of the inhibitor in which X = N-Me-Ala (figure1), were studied by carrying out molecular dynamics simulations, and energy minimizations.

COMPUTATIONAL PROCEDURES

We present here the different techniques used in this study:

Conformational search

Molecular dynamics simulations were performed at 600K in order to efficiently explore the conformational space of the molecule. The initial conformations were energy minimized by 200 steps of adopted basis Newton-Raphson and used as starting point for the molecular dynamic simulations. The temperature of 600K was fixed by increasing gradually the kinetic energy during 2.4 ps. It was followed by an equilibration stage lasting 3 ps. The final trajectories used for the analysis were realized for a period of 140 ps. The results reported here were based on coordinates averaged over 1ps time periods.

Contour map

The adiabatic energy contour map was built by constraining the two dihedral angles of interest around the grid point, by harmonic potentials. The conformer defined at each grid point, was energy minimized by using 200 steps of the adopted basis Newton-Raphson method followed by 5 steps of Newton-Raphson method.

Free energy difference

The thermodynamics integration method (1) was used to calculate the free energy difference between two conformers 1 and 2 defined by a single dihedral angle ω_0 . A driving potential was included in the potential energy in order to move this dihedral angle from its initial value ω_i to its final value ω_j (2). The driving potential was

$$V_\lambda = K(\omega - \omega_0)^2 \quad (K=10\text{kcal/mol}) \quad \text{with} \quad \omega_0 = (1-\lambda) \omega_i + \lambda \omega_j$$

where the reaction coordinate λ varies from 0 to 1. The free energy difference between the two conformers was expressed by

$$\Delta A_{1 \rightarrow 2} = \int_0^1 \left\langle \frac{\partial V}{\partial \lambda} \right\rangle_\lambda d\lambda$$

The quantities $\left\langle \frac{\partial V}{\partial \lambda} \right\rangle_\lambda$ were calculated for discrete values of λ in a series of dynamics carried out at 300K. The population ratio for the two conformers was estimated by the formulae:

$$\frac{P_2}{P_1} = e^{-\frac{\Delta A_{1 \rightarrow 2}}{KT}}$$

The program which we developed for the free energy difference calculation was included into the CHARMM program (3).

The model considered includes the solute surrounded by 367 water molecules. Conventional periodic boundary conditions were used. For each λ a trajectory of 15ps was realized, and only its last 4ps were used for the free energy estimation.

The calculations were performed using the program CHARMM(3). The extended atom model for incorporating the hydrogens into the heavy-atoms to which they are bonded was used only for aliphatic hydrogen atoms. The charges used are displayed on figure 1. The electrostatic energy between two atoms is calculated with a dielectric constant varying with the distance between these atoms in the vacuum calculations ; a dielectric constant equal to unity was used when water molecules were included into the model. These calculations were carried out on a microVAX and on the CRAY XMP at CEN Saclay.

RESULTS AND DISCUSSION

One of the major degrees of conformational flexibility of the molecule HS-CH₂-CH₂-CO-Pro-Ala comes from the internal rotation around the C α -CO bond of proline, (dihedral angle ψ_2 in figure 1).

It has been shown previously that in the case of a proline residue, the ψ_2 angle adopts mainly the three values : $\psi_2 = -60^\circ, 60^\circ$, and 150° corresponding respectively to

the classical most stable conformations CIS', C7 and TRANS'. Because the presence of a methyl group prevents the formation of the C7 conformer, the molecular dynamics simulations were performed starting only from TRANS' and CIS' conformers.

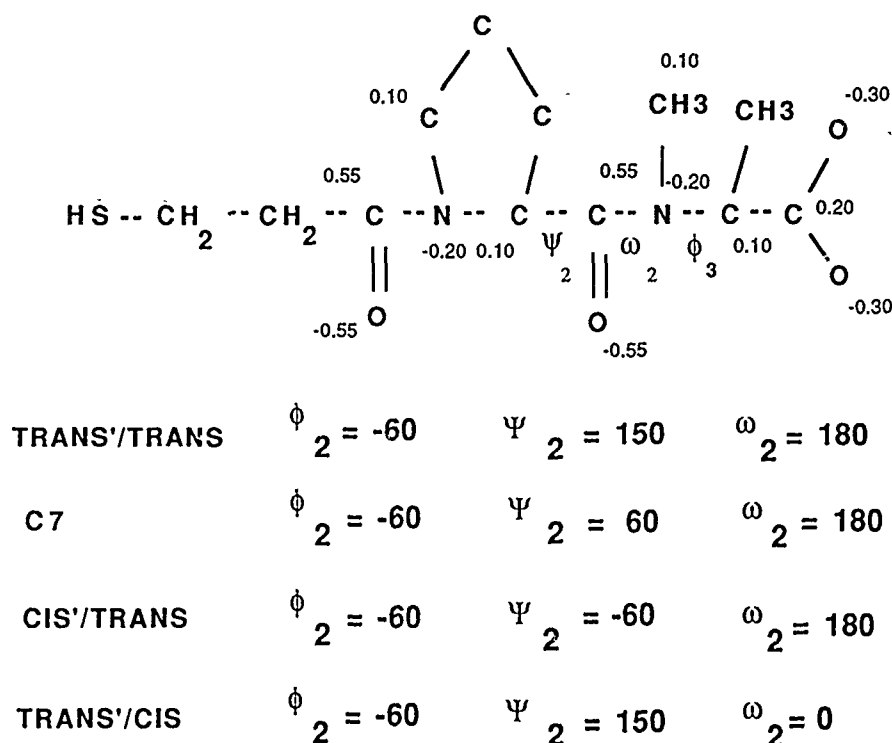
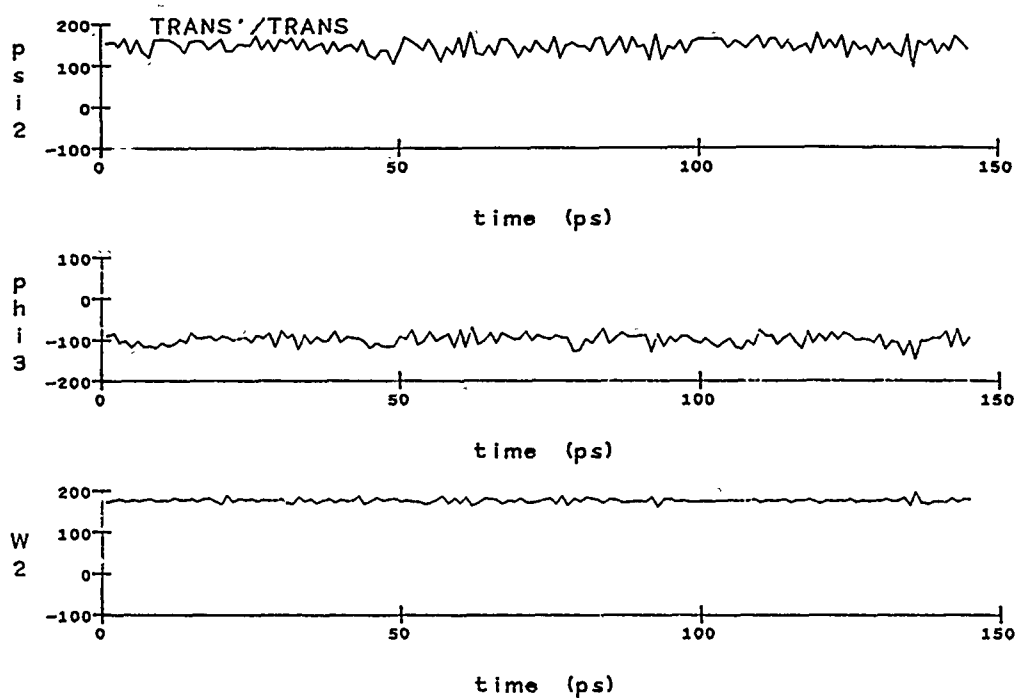
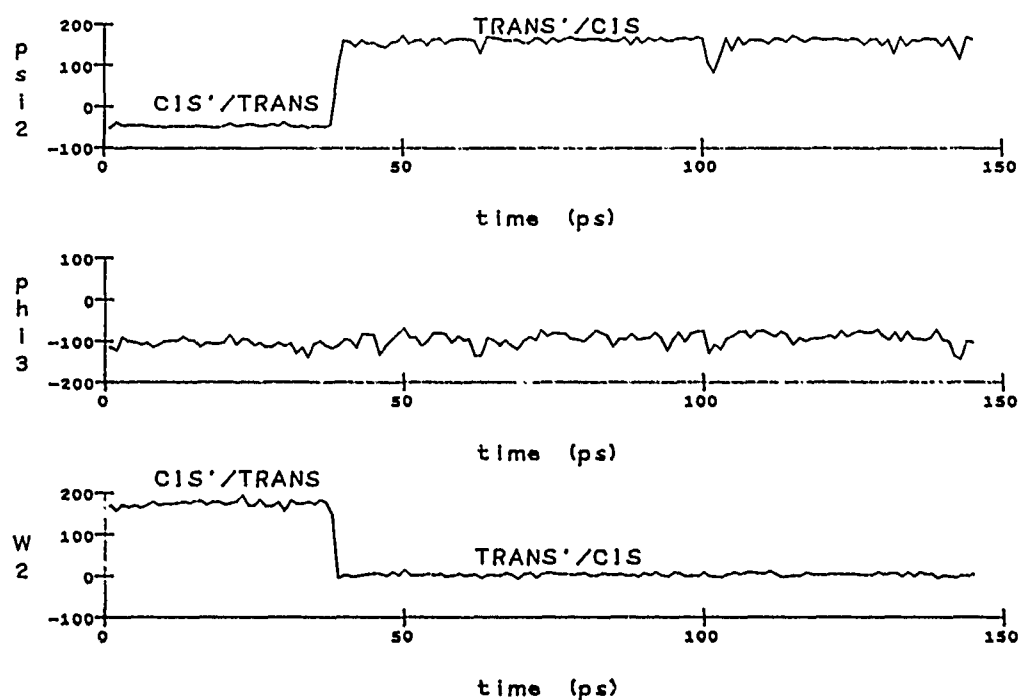


Figure 1 : N-Methyl Inhibitor bacterial collagenase (charges)

Figure 2a shows the values of the three dihedral angles ψ_2 , ϕ_3 , ω_2 as a function of time in the trajectory started from the TRANS' conformer. In this case no conformational change is observed during the time course of 140ps. In the contrary, figure 2b which corresponds to the trajectory started from the CIS' conformers shows a conformational change after a period of 40ps. The transition observed for ψ_2 (CIS' to TRANS') is coupled with that observed for ω_2 ; the last one corresponds to a transition from TRANS to CIS of the N-methylated peptide bond ($\omega_2 = 180^\circ$ to $\omega_2 = 0^\circ$).

This study was followed by the investigation of the energy minima corresponding to the TRANS'/TRANS and TRANS'/CIS conformers which were determined by the molecular dynamics simulation. Figure 3 shows the adiabatic contour map (ψ_2 versus ω_2) obtained by the procedure defined previously. The two conformations characterized by ($\psi_2 = 150^\circ$, $\omega_2 = 180^\circ$) and ($\psi_2 = 150^\circ$, $\omega_2 = 0^\circ$) correspond to two equal energy minima, whereas the CIS'/TRANS conformer corresponds to high energy minimum (18kcal/mol). These results may explain the fact that the molecular dynamics simulation started from

Figure 2a : values of ψ_2 , ϕ_3 , and w_2 from TRANS'/TRANSFigure 2b : values of ψ_2 , ϕ_3 and w_2 from CIS'/TRANS

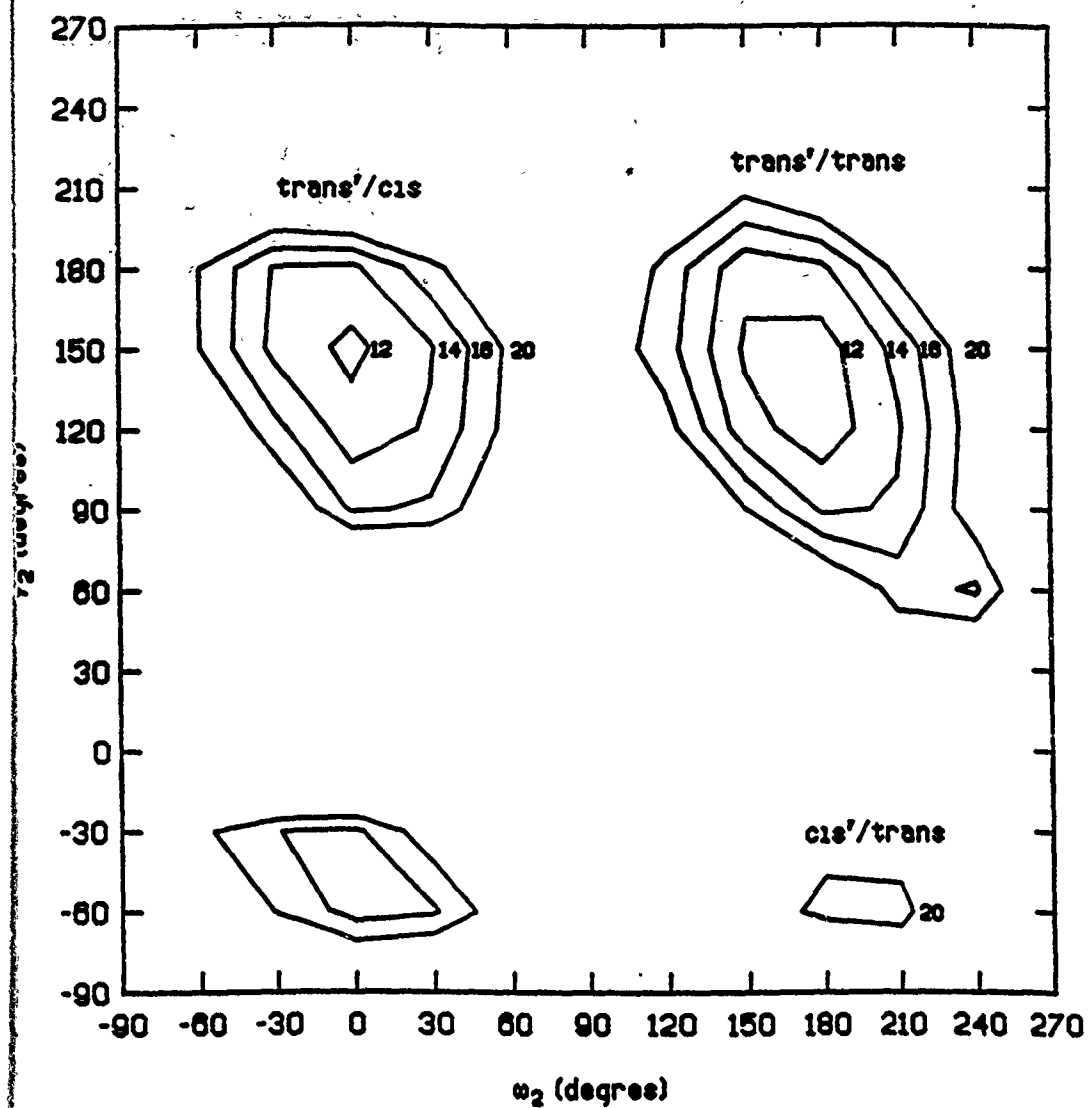


Figure 3 : Energy contour map for HS-CH₂-CH₂-CO-Pro-N-ME-ALA (in Kcal/Mol)

the CIS'/TRANS conformer converged to a conformer of lower energy, here TRANS'/CIS (figure 2b).

The conformers TRANS'/CIS and TRANS'/TRANS corresponding to the two lowest energy minima are the only one observed in the NMR experiments performed recently in our laboratory on the same compound. Moreover, these NMR experiments indicate that the population's ratio is 95% for the TRANS'/TRANS conformational state and 5% for the TRANS'/CIS one. This result is in accordance with the larger surface area observed for the TRANS'/TRANS conformer in comparison with that of TRANS'/CIS conformers in figure 2. In order to evaluate theoretically the population ratio between these 2 conformational states we performed free energy difference calculations.

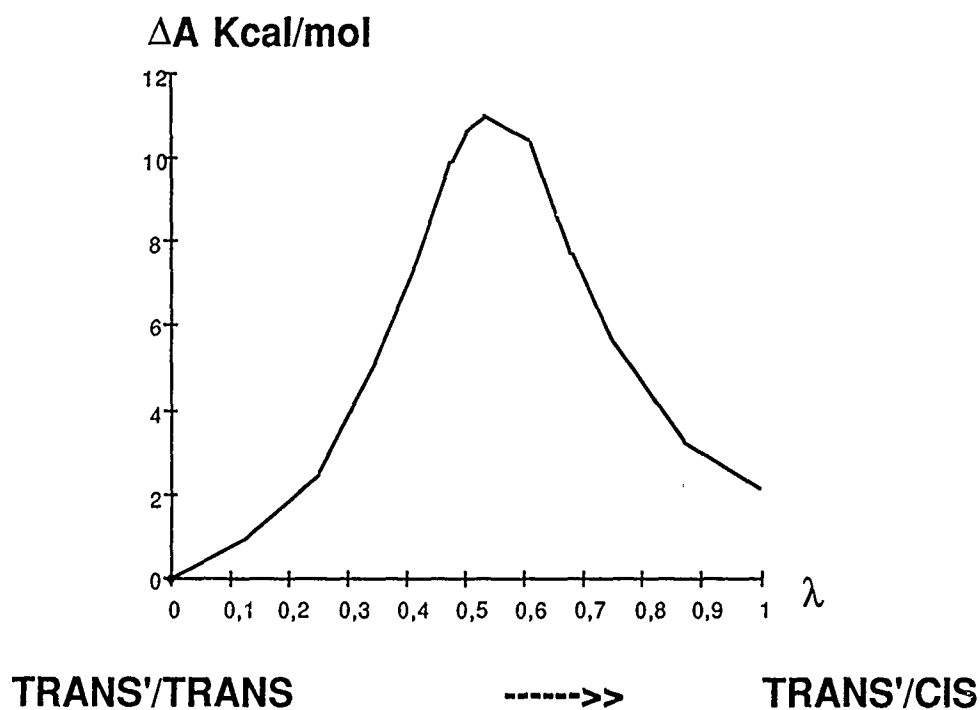


Figure 4 : Free energy for the transition TRANS'/TRANS to TRANS'/CIS

Figure 4 shows the variation of the free energy difference as a function of λ when going from TRANS'/TRANS to TRANS'/CIS conformation. The free energy difference between the two conformers is about 2 kcal/mol in favour of TRANS'/TRANS with a corresponding population of 96%. Although this result agrees well with the value given by the NMR results, longer simulations are needed to give an error estimation.

CONCLUSIONS

The calculations reported here show how molecular dynamics calculation are efficient to localize low energy minima. The consideration of the entropic factor appears here essential in order to explain the percentage of the different conformers observed experimentally for the N-methyl compound.

From the biological point of view, it appears that the stabilization of the TRANS'/TRANS conformation can explain the stronger inhibitor activity of the N-methyl compounds in comparison with that of the non methylated ones.

Considering the fact that a fast kinetics is observed for the inhibitory activity of this compound and dominance of this conformation in solution, and the large energy barrier around its energy minimum, it can be postulated that the TRANS' conformation corresponds to that of the bound state.

REFERENCES

- 1 Brooks III C.L., Karplus M., Pettitt B.M. Proteins : A Theoretical Perspective of Dynamics, Structure, and Thermodynamics Ad. in Chem. Phys. Ed: Wiley 1988
- 2 Anderson, A., Carson, M., Hermans, J. Ann. N.Y. Acad. Sci. 482 51-59 (1986)
- 3 Brooks B.R., Bruccoleri R.E., Olafson B.D., States D.J., Swaminathan S., Karplus M. J. Comp. Chem. 4, 187-217 (1983)

MOLECULAR DESIGN OF NEW BIOACTIVE MICROPROTEINS FROM THE KNOTTED STRUCTURE OF THE TRYPSIN INHIBITOR EETI II.

D. Le Nguyen, A. Heitz, L. Chiche, D. Nalis and B. Castro*

Centre CNRS-INSERM de Pharmaco-Endocrinologie, 34094 Montpellier France

*present address : Sanofi-Chimie 17 rue des Fossés St-Jacques, 75005 Paris France

A. Favel, H. Mattras and M.A. Coletti-Previero

INSERM U58, rue de Navacelles 34090 Montpellier France

C. Gaboriau and J.P. Mornon

Laboratoire de Cristallographie

Universite Pierre et Marie Curie, Place Jussieu, 75005 Paris France

Introduction

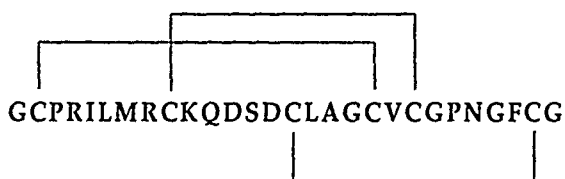


Figure 1 : The primary structure of EETI II

Seeds of *Ecballium Elaterium* (jumping cucumber) are a rich source of protein inhibitors of proteinases : we recently isolated, sequenced, and characterized the main trypsin inhibitor [1], which, following the international nomenclature [2], is designed as EETI II.

Its primary structure indicates that it belongs to the squash family [3], a recently discovered family of serine protease inhibitors [4] of small peptides rich in disulfide bridges. The chemically synthesized EETI II [5] is the shortest microprotein inhibiting a serine protease so far known, with an dissociation constant of $1.2 \cdot 10^{-11}$ M. It contains 28 amino acid residues and three disulfide bridges ; the primary structure is outlined on Figure 1.

Solution structure by NMR

Though the primary structure is closely related to other microproteins extracted from various cucurbitaceae, its chemical synthesis was facilitated by an unique feature of the C-terminal sequence, where a -Gly-22-Pro-23-Asn-24-Gly-25- β -turn forming section behaves as a strong template to drive the correct folding in high yield and selectivity. Large quantities of EETI II could be prepared, enough to make a complete 2D NMR study [6] ; this study allowed to depict the different secondary structure elements of EETI II ; however, the access to the disulfide bonds connectivity remained unclear. Indeed this assignment could not proceed even on the simple examination of the nOe's displayed between the cystein residues ; Cys2 was completely lacking of long range nOe's, Cys21 had two nOe's with both Cys9 and Cys27 : on the other hand Cys15 exchanged a nOe with Cys 19 (see Figure 2).

Hence a DISGEO modelisation of all 15 possible disulfide bridges combinations made

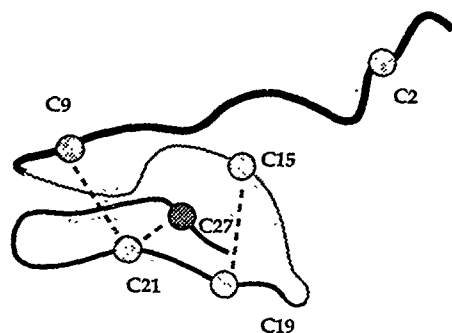


Figure 2 : nOe's involving cysteines in EETI II

possible the assignment of the disulfide bridges : 4 structures were characterized by a minimal number of distance violations ; in these structures it appeared that the main difference lied in the conformation of the Gly1-Arg8 strand, governed by the connexion imposed to Cys2. These structures are named from S1 to S4 with increasing number of distance violations and outlined on Figure 3 together with the values of the torsion angles of the disulfide bridges ; the rejection of structures S3 and S4 was dictated by the unacceptable value of the dihedral angle of the Cys21-Cys27 bridge ; structure S2 was rejected because of the

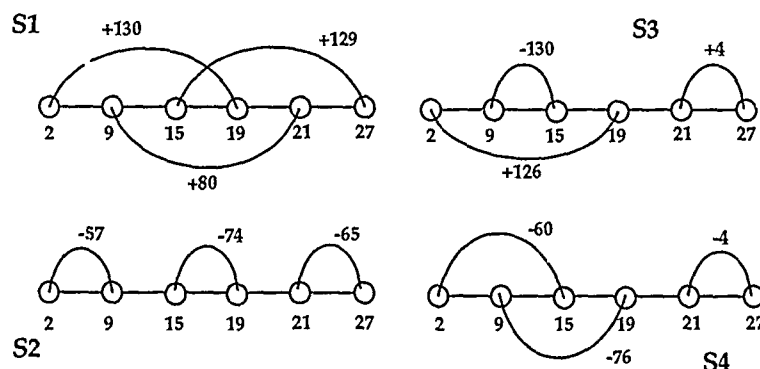


Figure 3 : disulfide bonds connectivities and torsion angles for the 4 best DISGEO structures of EETI II

lack of interlacing of the disulfide bridges, unable to justify the apparent rigidity of the molecule.

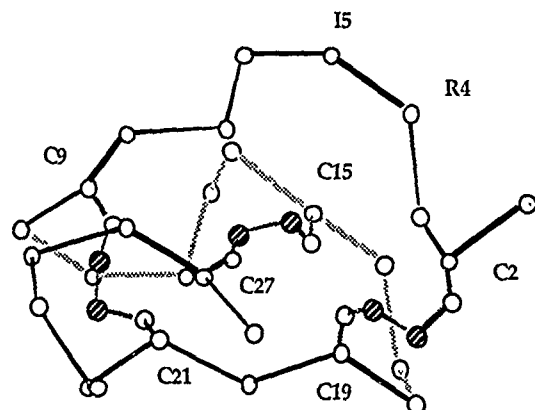


Figure 4 : DISGEO model of EETI II ; C α chain and disulfide bridges.

in the crystal complex, though the carbonyl bond of Arg 4 lies in the vicinity (2.4 Å) of the distal nitrogen atom of His42 of the enzyme. The interaction of the inhibitor with the

A very striking characteristics of this molecule lies in the fact that the Cys15-Cys27 bridge penetrates the macrocycle formed by the two other disulfide bridges and the strands linking them. This fact is illustrated on Figure 4 .

EETI / trypsin crystal structure.

The radiocrystallographic studies of the parent Cucurbita maxima / bovine trypsin inhibitor (CMTI I)[7] and by our own crystal structure of the EETI II / porcine pepsin structure [8] have been made recently.

Crystal structures confirm unambiguously our NMR assignments of the disulfide bridges. It appears that the sessile Arg4-Ile5 bond is not cleaved in the

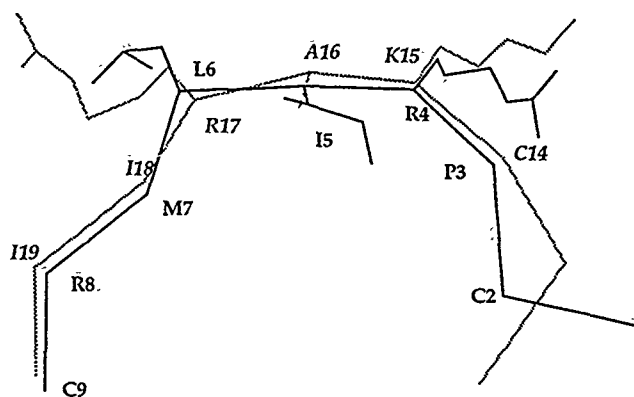


Figure 5 : superimposition of the inhibiting loops of EETI II (black line, roman) and BPTI (gray line italic)

enzyme involves the strand running from Gly1 to Leu6 occupying the sites from P2 to P'4. Strikingly, Leu6 shares the interaction in P'2 (Tyr141) with Phe26. Another interesting feature lies in the close similarity between the inhibiting-loop of EETI II and the corresponding strand of BPTI. C α from 3 to 8 of EETI II can be fitted with C α from 14 to 19 of BPTI, in spite of the lack of similarity in the sequences as shown on Figure 5. This result supports the hypothesis of the

existence of a variable sequence, fixed conformation loop in a large superfamily of serine proteinase inhibitors [9].

EETI II as a vehicle for new functions.

We have found that EETI II was a unique carrier for new inhibiting functions ; active site modifications can produce inhibiting activities orientated towards other serine proteinases ; furthermore, we succeeded in grafting a second active site inhibiting carboxypeptidase A.

Elastase inhibitors.

Elastases are serine proteinases homologous to trypsins ; the biological role of human leucocyte elastase in the microbial lysis by polymorphonuclears has been cleared up. However this powerful and poorly selective proteinase is normally completely inhibited by the α 1-antitrypsin (a misnamed circulating protein that should be called anti-elastase, synthesized in the hepatocyte). Defective secretion of α 1-antitrypsin results in a pathogeneus activity of elastase, especially directed against lung elastin, a constitutive protein of lung tissues ; those people presenting such disease generally suffer severe chronic emphysema. It is well known that the substrate specificity of elastases differs from that of trypsin in the nature of the residue placed in S1 recognition subsite. Trypsin prefers a basic, positively charged residue such a lysine or an arginine, porcine pancreatic elastase prefers a neutral small hydrophobic alanine, whereas human leucocyte elastase best substrates bear an hydrophobic, bulky valine. The simple substitution of Arg4 of EETI II by alanine and valine induced both the disappearance of the

	Trypsin	HLE*	PPE**	α -Chymotrypsin
EETI II [1-28]	1.10^{-12} M	-	-	-
Ala-4	-	$1.4 \cdot 10^{-8}$ M	$2.5 \cdot 10^{-8}$ M	-
Val-4, Nle-7	-	$2.2 \cdot 10^{-9}$ M	$5 \cdot 10^{-6}$ M	-
Phe-4, Nle-7	-	-	-	$2 \cdot 10^{-7}$ M

Table : Dissociation constants of EETI analogues towards three serine proteinases (* human leucocyte elastase ; **porcine pancreatic elastase).

anti-trypsin activity and the promotion of a very strong inhibition of pancreatic and leucocyte elastase. Similarly the replacement of Arg4 by a phenylalanine promotes an anti-chymotrypsin activity as shown in the Table [10].

A new family : the knotted proteins.

The striking knotted structure where the disulfide bridges between the third and the sixth cysteines crosses the macrocycle formed by the first, second, fourth and fifth cysteines was also found in CPI, a carboxypeptidase inhibitor found in potato leaves [11]. Conotoxin- ω [12] appeared recently as a member of this new family [13]. These three microproteins share the same 6 cysteines network [I-IV, II-V, III-VI] with very few sequence homology, and constitute a new structural family of proteins sharing the same

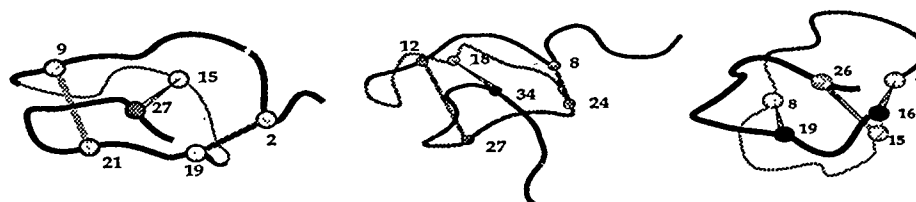


Figure 6 : the knottins ; left : EETI II ; middle : CPI ; right : conotoxin ω

topology, the "knottins" (Figure 6) ; the cysteine interlocked connectivity is necessary but not enough to insure the belonging to this family ; indeed, each of the four domains of wheat germ agglutinin have such a connectivity but with a planar spiral topology [14]. The recently studied E. coli enterotoxin [15] appeared to be in the same case of an interlocked but topologically planar molecule [13]. It can be seen on Figure 7 that very few, if no sequence homology is shared by these three molecules.

The folding mechanism of such molecules is especially intriguing and EETI II shall certainly prove to be an excellent model for such a study.

	G	C	P	R	I	L	M	R	C	K	Q	-	D	S	D	C	L	-	A	G	-	C	V	-	C	G	P	N	G	F	-	-	C	G	EETI II				
<EQHADPI	C	N	K	P	-	-	-	-	C	K	T	H	D	-	D	C	S	G	A	W	F	C	Q	A	C	W	-	N	S	A	R	T	C	G	P	Y	V	G	CPI
	G	C	P	R	I	L	M	R	C	K	Q	-	D	S	D	C	L	-	A	V	-	C	V	-	C	G	P	N	G	F	-	-	C	G	P	Y	V	G	TCPI*
	C	K	S	X	G	S	S	C	S	X	T	S	Y	N	C	-	-	-	-	-	-	C	R	S	C	N	X	Y	T	K	R	-	C	H	-	n	h	2	conotoxin ω

Figure 7 : sequence alignment of the knottins. (*see below)

TCPI, a chimeric microprotein inhibitor of both trypsin and carboxypeptidase A

These considerations prompted us to synthesize a "chimeric" peptide with the sequence of EETI II plus the C-terminal tetrapeptide of CPI in order to check to possibility of building a double headed inhibitor directed towards two enzymes vastly different in their active sites.

Figure 7 shows the alignment of the sequences of EETI II and CPI, fixed on the six cystein residues. Apart the cysteines, only six residues are shared by both inhibitors. The C-terminal inhibiting tail of the CPI, with Tyr, Val and Gly in position P2, P1 and P'1

of carboxypeptidase is completely lacking in EETI II ; on the other hand, the inhibiting loop of the trypsin inhibitor, Pro-Arg-Ile-Leu-Met-Arg is shortened to three residues in the homologous segment of CPI.

Fitting the coordinates of the six cysteines of each inhibitor/enzyme complex on a computer graphic screen shows that it would be conceivable to make an inhibitor binding both enzymes when the two binding sites are present: indeed, as these binding sites are apart from the general plane of the molecule, the two enzymes can stay without any major steric interaction around a tentative model of the extended EETI II molecule (TCPI) as seen in Figure 8.

Indeed, the extended EETI II shows stoichiometric interaction and good inhibitory potency (Trypsin : $K_d = 1.8 \cdot 10^{-9}$ M ; CPA : $K_d = 3 \cdot 10^{-9}$ M) towards its target enzymes measured separately. However, the inhibitory constant towards trypsin is significantly

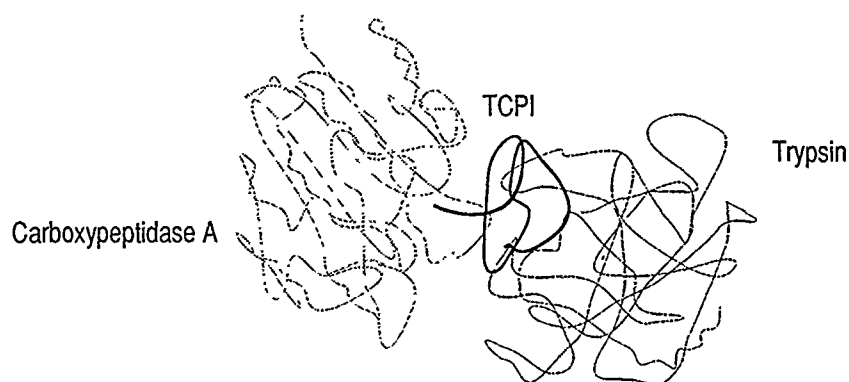


Figure 8 : Predictive model of the CPA / TCPI / trypsin complex

lower than in the original EETI II ; this loss is attributed to the extended C-terminal which may impair the ionic interaction of Gly-28 with the N-terminus[7]. By contrast, the inhibition constant towards CPA is similar if not identical to that of the original potato inhibitor, thus showing both the small influence on inhibitory activity of the primary sequence of the N-terminal part of this molecule and the importance of the overall topology due to the arrangement of the disulfide bridges.

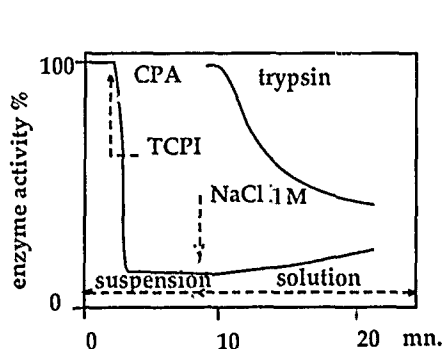


Figure 9 : time course of the CPA / TCPI / trypsin complex formation

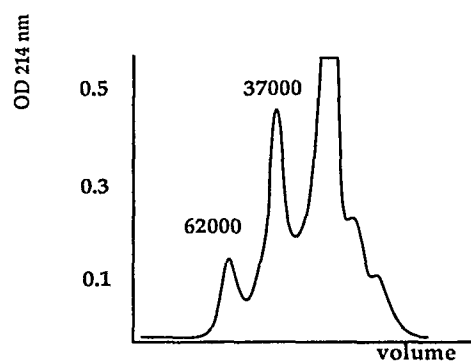


Figure 10 : molecular sieve elution profile of the trimolecular complex

In order to ascertain whether the bis-headed peptide was able to react concomitantly with both enzymes ; the 1/1 insoluble CPA/peptide complex was solubilized in 1 M NaCl and allowed to interact with trypsin in less than stoichiometric quantity. Both activities were reduced as shown in Figure 9 and the molecular sieve chromatography of the reaction mixture showed a peak in the molecular weight range of 62000, which could be the tri-component complex (Figure 10 ; M.W. = 61860 for trypsin/peptide/CPA) [16].

Nothing could have allowed such a synthesis beside molecular modelling. Neither the origin of the two peptides, (one from the Cucurbitaceae, the other from the Solanaceae), nor their function, (one inhibits a serine-protease the other a zinc protease), nor the sequence (beside the cysteine positions, very little similarity is shown). Only the knowledge of the 3D structure with the knotted disposition of the disulfide bridges enlightened the similarity and suggested the graft of an added active site to an already active, albeit different peptide.

References

1. Favel, A., Mattras, H., Coletti-Previero, M.-A., Zwilling, R., Robinson, E.A. and B. Castro (1988) *Int. J. Peptide Protein Res.* 33, 202-208.
2. Wilusz, T., Wiecek, M., Polanowski, A., Denton, A., Cook, J. and Laskowski, M. Jr. (1983) *Hoppe-Seyler's Z. Physiol. Chem.* 364, 93-95.
3. Wiecek, M., Otlewski, O., Cook, J., Parks, K., Leluk, J., Wilimowska-Pelc, A., Polanowski, A., Wilusz, T. and Laskowski, M. Jr. (1985) *Biochem. Biophys. Res. Com.*, 126, 646-652.
4. Powers, J.C. and Harper, W. (1986) in *Proteinase Inhibitors*, Barrett, A.J. and Salvesen, G. Ed. Elsevier, 55-152.
5. Le-Nguyen, D., Nalis, D., Heitz, A. and Castro, B. (1989), *Int. J. Peptide Protein Res.* in press.
6. Heitz, A., Chiche, L., Le-Nguyen, D. and Castro, B. (1989), *Biochemistry*, 28, 2392-2398.
7. Bode, W., Greyling, H.J., Huber, R., J. Otlewski, J. and Wilusz, T. (1989), *FEBS Lett.* , 242, 285-292.
8. Gaboriau, C., Vaney, M.C., Bachet, B., Le-Nguyen, D., Castro, B. and Mornon, J.P., *J. Mol. Biol.* (1989), in press.
9. Kelly, C., Otlewski, J., Qasim, M.A., and Laskowski, Jr. M., (1989), Abstract presented at the 3rd Symposium of the Protein Society, Seattle.
10. Favel, A., Le-Nguyen, D., Coletti-Previero, M.A., and Castro, B. (1988), *Biochem. Biophys. Res. Comm.* 162, 79-82.
11. Rees, D.C. and Lipscomb, W.N. *J. Mol. Biol.* (1982), 160, 475-498.
12. Nishiuchi, Y., Kumagaye, K., Noda, Y., Watanabe, T.X. and Sakakibara, S. *Biopolymers*, (1986), 25, S61-S68.
13. Kobayashi, Y., (1989), personal communication.
14. Wright, C.S., (1981), *J. Mol. Biol.*, 145, 453-461.
15. Gariépy, J., Judd, A.K., and Schoolnik, G.K. (1987), *Proc. Natl. Acad. Sci. USA*, 84, 8907-8911.
16. Le-Nguyen, D., Mattras, H., Coletti-Previero, M.-A., and Castro, B. *Biochem. Biophys. Res. Comm.* (1989) 162, p. 1425-1430.

INTERMOLECULAR FORCES IN SOME IONIC ATMOSPHERIC CLUSTERS INVOLVING WATER. QUANTUM MECHANICAL AND MONTE CARLO CALCULATIONS

E. KOCHANSKI ¹, A. RAHMOUNI ¹, R. WIEST ²

¹ UPR n°139 du CNRS, Laboratoire de Chimie Théorique,

² UPR n°139 du CNRS, Laboratoire de Chimie Quantique,
Université Louis Pasteur, Institut de Chimie, BP 296, 67008 Strasbourg-Cedex, France

SUMMARY

On the example of the system $\text{H}_3\text{O}^+(\text{H}_2\text{O})_n$, the effect of the energy parameters on Monte Carlo clustering energies of small clusters is studied. Special attention is paid to three-body contribution. Quantum mechanical calculations (ab initio SCF and dispersion energies) are performed on some specific geometrical structures of cluster $\text{H}_3\text{O}^+(\text{H}_2\text{O})_6$, allowing to study three and higher-body effect on the first solvation shell. Calculations are also done with some of the most known expressions and parameters available in the literature.

INTRODUCTION

Monte Carlo and Molecular Dynamics calculations on systems bonded through intermolecular interactions are generally based on the use of analytical expressions describing the interaction between two, or eventually three, molecules. In most cases, the energy of the total system is computed as a sum of intermolecular energies between pairs of molecules, within the "pair approximation", generally assuming pairwise additivity. Corresponding parameters are needed. They may be adjusted either to reproduce experimental data or to fit previous calculations of energies and properties of specific geometrical configurations of the bimolecular systems. In the first case, the data available generally involve three and higher-body contributions, which leads to introduce some artefact in the determination of the parameters; we then have "effective" potentials : it is assumed that three and higher-body terms are partly taken into account through two-body expressions.

Many expressions and parameters have been proposed for the water system (refs. 1-7). Comparisons are available, but they generally concern large systems, considering averaged values, and it is difficult to really know their validity in the case of specific geometries. In the present work, quantum mechanical calculations are presented for four geometries of clusters involving six water molecules. These energies are compared with those obtained from some of the most known expressions and parameters available in the literature.

Some parameters describing the interaction between H_3O^+ and a water molecule have been proposed previously (refs. 8-11). In the present work, a comparison is done for some specific geometries, considering ab initio SCF and dispersion energy calculations together with some expressions available in the literature.

Three-body forces have also deserved special attention. Their effect on Monte Carlo aggregation energies (ref. 11) and in quantum mechanical calculations performed on some of the most stable configurations of small clusters (ref. 12) have been shown previously. In the present work, this effect is also studied from quantum mechanical calculations performed on some specific geometries of clusters involving 6 water molecules. This gives interesting information on the nature of the filling of the first solvation shell. As will be seen, the neglect of three-body contributions can artificially favor some geometries.

DETAIL OF THE CALCULATIONS

When **Monte Carlo results** are mentioned, we refer to calculations based (refs. 13-14) on the use of an atom-atom (12-1-6) expression to describe the interaction between the ion and each water molecule, the water-water interaction being evaluated from **MCY** potentials (ref. 4). The effect of three-body forces, described by an atom-atom expression, has also been investigated (refs. 11-12). **Two geometries** of the **oxonium ion** have been considered : a planar one, **PL**, with $d(\text{OH}) = 0.965 \text{ \AA}$ and $\theta = 120^\circ$ (ref. 9), and a slightly pyramidal one, **PYR**, with $d(\text{OH}) = 0.959 \text{ \AA}$ and $\theta = 113.5^\circ$ (ref. 15). Two sets of parameters have been used for the ion-water interaction, denoted by (12-1)(6) and (12-1-6), respectively : in the first case, the parameters have been adjusted to represent separately the SCF and dispersion energies computed for some geometrical configurations of the bimolecular system; in the second case, the parameters are adjusted to reproduce the total energies (these procedures are commented in refs. 11 and 16).

One of the important feature of Monte Carlo calculations based on the pair approximation is the possibility to find, among the most stable configurations generated, either 3 or 4 water molecules in the first solvation shell when the total number of water molecules in the cluster is at least 6. Special attention is then paid to such configurations. **SCF ab initio calculations** have been performed (using a vectorized version of program **ASTERIX** developed in Strasbourg) with the two geometries of the ion and structures involving 3 or 4 water molecules in the first shell of clusters $\text{H}_3\text{O}^+(\text{H}_2\text{O})_6$. These structures are denoted by (3+3)PL, (4+2)PL, (3+3)PYR, (4+2)PYR, the notation (4+2) meaning that 4 water molecules are in the first shell and the other 2 in outer shells. **DZP** basis sets are used (ref. 9), and the SCF intermolecular values are corrected for the basis set superposition error (BSSE), using the counterpoise correction (ref. 17). In order to evaluate the **three and higher-body contribution** effect, **TB**, SCF calculations have been performed on the total supermolecule system (**SCF(SM)**), and for all pairs of molecules (**SCF(bi)**). The **dispersion energy** is derived from our previous approximation **EK** (ref. 9), and from the expansion **RW** proposed by Wormer et al (ref. 18) on the basis of accurate ab initio quantum mechanical calculations in the perturbation scheme and the multipole expansion. More details on this part of the work are given in ref. 19.

Calculations have also been performed on the 4 structures described above, using some expressions and parameters available in the literature. They are denoted by **EK**, **GGB**, **BB** in the case of ion-water interaction, corresponding to references (9,10,8), respectively. For the widely studied water-water system, we generally refer to

the usual notation when available and to the initials of the authors otherwise. Parameters: MCY (ref. 4), AFHF (ref. 3), SPC (ref. 5), BF (ref. 1), TIP3P (ref. 7), TIPS2 (ref. 7), TIP4P (ref. 7), ST2 (ref. 2), CH (ref. 6), and RW (ref. 18) have been considered. When the dispersion energy is taken from different references for ion-water and water-water pairs, we use the notation dispIW(XX) or dispWW(YY) , XX or YY referring to the expressions and parameters described above. Let us note that RW parameters are directly computed from quantum mechanical ab initio calculations; EK, MCY, AFHF and CH parameters are adjusted to reproduce ab initio calculations of surfaces, while experimental data are taken into account in the other cases, introducing semi-empirical considerations.

THE RESULTS

Three-body forces effect on the clustering energies

Table 1 shows the effect of the energy parameters on Monte Carlo clustering energies of small clusters of the system $\text{H}_3\text{O}^+(\text{H}_2\text{O})_n$, special attention being paid to three-body contribution. The two series of parameters (12-1)(6) and (12-1-6) and the two geometries of the ion (planar or pyramidal) described above are considered (ref. 11).

Our previous values (ref. 20) were obtained with a (12-1)(6) fitting but different $C_{ij,6}$ parameters. We can see that the use of parameters (12-1)(6) or (12-1-6) in the pair approximation involves only minor changes within the pair approximation, both geometries giving similar qualitative results. But three-body forces (column noted 3B) strongly affect the clustering energies of small clusters, giving values in much better agreement with experimental data. The result for $n = 2$ remains larger than the experimental one. It would be interesting to check if tunneling process could be responsible for this effect. As seen in ref. 11, three-body effects seem much less important in larger systems.

Three-body effect is particularly sensitive on this system. It is much less pronounced with calcium ionic clusters (ref. 16).

TABLE 1

$\text{H}_3\text{O}^+(\text{H}_2\text{O})_n$: Monte Carlo clustering energies $E_n - E_{n-1}$ in Kcal/mol. $T = 300 \text{ K}$

n	Planar				Pyramidal		Experimental	
	(12-1)(6)	(12-1)(6)	(12-1-6)	Ref.(20)	(12-1)(6)	(12-1-6)	Ref.(21)	Ref.(22)
	3B							
1	-30.08		-29.17	-29.05	-32.41	-30.85	-31.6	-31.8
2	-28.72	-24.41	-27.83	-27.58	-30.71	-29.34	-19.5	-19.0
3	-27.26	-19.68	-26.38	-26.25	-28.94	-27.69	-17.9	-17.6

Three-body forces effect on the filling of the first solvation shell

We previously noted (ref. 20) that the most stable configurations generated in Monte Carlo calculations based on the pair approximation have 3 water molecules in the first solvation shell till $n = 5$, as could be expected from the formation of hydrogen bonds with the three H atoms of the ion. However, from $n = 6$, we found configurations

of similar energies with either 3 or 4 water molecules in the first shell. We proposed (réf. 20) that this fourth water molecule can be explained by an exchange process through one of the three hydrogen bonds. It is interesting to study three-body effect on such structures. As described above, and with more details in ref. 19, ab initio SCF calculations have been performed on structures (3+3)PL, (4+2)PL, (3+3)PYR, (4+2)PYR of the system $\text{H}_3\text{O}^+(\text{H}_2\text{O})_6$. Table 2 shows the three and higher-body contribution in the whole system, TB, from the difference between SCF(SM) and SCF(bi). We can see that this contribution, rather similar for both geometries of the ion, is clearly different for the (3+3) and (4+2) structures, so that the four structures have close energies in a pair approximation (SCF(bi)) while the (3+3) geometries are more stable than the (4+2) ones by 7-9 kcal/mole in the supermolecule treatment (SCF(SM)). Since three-body contribution occurs mainly at SCF level, it is clear that the neglect of this effect favors the (4+2) structures.

TABLE 2
 $\text{H}_3\text{O}^+(\text{H}_2\text{O})_6$: SCF energies

	3+3 PL	4+2 PL	3+3 PYR	4+2 PYR
Total system				
SCF(SM)	-96.9	-89.3	-95.2	-86.8
SCF(bi)	-105.6	-106.6	-105.1	-103.8
TB	8.7	17.3	9.8	17.0
Separated systems				
\sum I-W : SCF(bi)	-102.2	-113.1	-95.1	-106.7
\sum I-W : SCF(bi) + TB	-93.5	-95.7	-85.2	-89.7
\sum W-W : SCF(bi)	-3.3	6.4	-10.0	2.9

We may assume that this many-body contribution is mainly due to ion-water-water interaction, and decompose the total energy into ion-water molecules (\sum I-W) and water-water molecules (\sum W-W) components. At SCF level, we can see that the ion-water molecules contribution always favors the (4+2) structures, even when three and higher-body terms are taken into account, while the water-water molecules component favors the (3+3) ones. It is then clear that a correct description of the total system requires a good accuracy on both components.

Comparison of expressions and parameters : $\text{H}_3\text{O}^+(\text{H}_2\text{O})_6$ structures

Many expressions and parameters have been proposed in the literature for the water-water energy. At SCF level (we call "SCF level" any expression which does not include the dispersion term), Table 3 presents a wide variety of values. However, in all cases, the most stable structure is (3+3)PYR, followed by (3+3)PL, the other two being generally repulsive, except in ST2 calculations. But we must notice that the repulsive energy in the (4+2)PYR structure varies from 2.9 to 13.7 kcal/mole. If we consider the comments mentioned above about the required accuracy, this situation is rather worrying.

TABLE 3

 $\text{H}_3\text{O}^+(\text{H}_2\text{O})_6$: Water-Water Molecules Contribution, SCF level (no dispersion energy) ^a

$\Sigma(\text{W-W})$	3+3 PL	4+2 PL	3+3 PYR	4+2 PYR
SCF (bi)	-3.3 ^b	6.4	-10.0	2.9
SPC	-1.6 ^b	12.1	-6.2	13.7
	-4.2 ^c	9.6	-10.1	10.2
BF	-3.7 ^b	7.2	-8.5	7.4
	-4.0 ^c	6.8	-9.0	6.8
TIP3P	-2.6 ^{b,c}	11.3	-7.6	12.4
TIPS2	-4.0 ^{b,c}	9.1	-9.6	9.5
TIP4P	-4.5 ^{b,c}	7.6	-10.0	7.5
ST2	-8.3 ^c	1.1	-16.0	-4.6

a) See the text for the notations

b) molecular geometries described in the present work

c) molecular geometries described in the corresponding reference

TABLE 4

 $\text{H}_3\text{O}^+(\text{H}_2\text{O})_6$: Water-Water Molecules Contribution, Dispersion Energy ^a

$\Sigma(\text{W-W})$	3+3 PL	4+2 PL	3+3 PYR	4+2 PYR
SPC	-4.9	-6.8	-7.3	-9.2
BF	-6.6	-9.1	-9.7	-12.4
TIP3P	-4.7	-6.5	-6.9	-8.8
TIPS2	-4.7	-6.5	-7.0	-8.9
TIP4P	-4.8	-6.6	-7.1	-9.0
ST2	-2.1	-2.9	-3.1	-4.0
RW ^d	4.6 ^e	-6.8	-6.9	-9.1
	-5.7 ^f	-8.3	-8.4	-11.2
	-11.3 ^g	-20.8	-16.9	-29.5

d) ref. 19

e) TDCHF : $C_6 R^{-6}$ termf) MBPT : $C_6 R^{-6}$ termg) TDCHF : $C_6 R^{-6}$ to $C_{10} R^{-10}$ terms

The dispersion energies presented in Table 4 correspond to the leading term $C_6 R^{-6}$, except for the last line of the RW development (ref. 19) which includes terms till R^{-10} . All semi empirical determinations are generally quite close together, though BF results are somewhat larger and ST2 values somewhat smaller. We may note that these values are very similar to those obtained from the quantum mechanical development of Rijks and Wormer using the TDCHF approximation, while the MBPT values are larger by more than 20%. However, when the higher order terms are considered (last line of Table 5), the results may increase by a factor of 3. These last values are overestimated, due to the neglect of the overlap effect in this region, but we can expect that the true

results should be, in any case, larger than the values generally proposed on the basis of the leading term without overlap correction: the overestimation due to the charge overlap effect is most probably smaller than the higher order terms contribution. Combined with the remarks about the SCF level contribution (Table 3), we may question the accuracy of such calculations.

For the shortest water-water intermolecular distances, we noted (ref. 19) that the $C_{10} R^{-10}$ term becomes slightly dominant in the RW expansion. This can explain that hydrogen bonds are sometimes described by a (12-10) expression (refs. 23). However, we must keep in mind that such an expression may be eventually used only for very specific intermolecular (or interatomic) distances. In fact, the contribution of the other terms is artificially taken into account through a modified C_{10} coefficient. The values given by such an expression for other intermolecular distances could be very bad.

The analysis of Tables 3 and 4 shows how delicate it is to obtain accurate values of the energies. From the total energy (Table 5), it is extremely difficult to decide which expansion is the best. We may note that, in all cases, the most stable structure is (3+3)PYR, but structures (4+2) have either attractive or repulsive energies.

TABLE 5
 $H_3O^+(H_2O)_6$: Water-Water Molecules Energy ^a

$\sum(W-W)$	3+3 PL	4+2 PL	3+3 PYR	4+2 PYR
SCF (bi) + disp(RW) ^d	-14.7	-14.4	-26.9	-26.6
SCF (bi) + disp(SPC)	-8.3	-0.4	-17.2	-6.4
SPC	-6.6 ^b	5.3	-13.5	4.4
	-9.1 ^c	2.8	-17.3	1.0
MCY	-10.6	1.3	-19.8	-6.7
CH	-13.7	-6.1	-25.5	-14.0
AFHF	-5.8	6.8	-12.7	1.9
BF	-10.3 ^b	-1.9	-18.2	-5.0
	-10.6 ^c	-2.3	-18.7	-5.6
TIP3P	-7.3 ^{b,c}	4.8	-14.5	3.6
TIPS2	-8.8 ^{b,c}	2.6	-16.6	0.7
TIP4P	-9.4 ^{b,c}	0.9	-17.1	-1.5
ST2	-10.4 ^c	-1.8	-19.1	-8.6

a-b-c): see footnote on Table 3;

d) ref. 19

A similar situation is found for the few expressions which describe the interaction between the ion and the water molecules. This is shown in Table 6, where we can also note the importance of the molecular geometry. In particular, the use of parameters EK, adjusted for a planar geometry of the ion, is not very suitable for a pyramidal geometry. This points out the difficulty to search the best geometry with a same set of parameters.

The discrepancies in the dispersion energy values are rather surprising. With such differences, we may question the real significance of these contributions and we

would rather suggest that only the total value of the energy should be considered with GGB parameters.

TABLE 6

$\text{H}_3\text{O}^+(\text{H}_2\text{O})_6$: Ion-Water Molecules Contribution ^a

$\sum(I-W)$	3+3 PL	4+2 PL	3+3 PYR	4+2 PYR
SCF level				
SCF(bi) + TB	-93.5 ^b	-95.7	-85.2	-89.7
SCF (bi)	-102.2 ^b	-113.1	-95.1	-106.7
EK	-109.4 ^b	-123.5	-116.1	-128.8
GGB	-70.1 ^b	-74.2	-66.7	-69.7
	-78.5 ^c	-81.1	-74.3	-76.9
Dispersion energy				
EK	-4.6 ^b	-4.6	-4.5	-4.8
GGB	-18.1 ^{b,c}	-24.2	-19.0	-25.2
Total energy				
EK	-114.0 ^b	-128.2	-120.7	-133.6
GGB	-88.2 ^b	-98.3	-85.7	-94.9
	-96.6 ^c	-105.3	-93.3	-102.0
BB	-85.7 ^b	-96.1	-85.9	-96.9
	-87.3 ^c	-97.0		
SCF(bi) + disp(EK)	-106.9 ^b	-117.7	-99.6	-111.5
SCF(bi) + disp(EK) + TB	-98.1 ^b	-100.3	-89.8	-94.5

a-b-c) : see footnotes on Table 3

From all these considerations, it is not surprising that no definitive answer can be given, at this stage of the work, about the relative stabilities of the 4 structures studied. Indeed, they are rather close in energy and still more accurate calculations should be necessary to remove the ambiguity. It is clear that the pair approximation favors the (4+2) structures (Tables 2, 7); however, the dispersion energy in water-water systems counterbalances three-body effect since it favors the (4+2) configurations by about 2 kcal/mole when the only leading $C_6 R^{-6}$ is considered, and by 9-13 kcal/mole for an expansion till the R^{-10} terms in the RW treatment (Table 4). In the last case, the total energies are (Table 7, SCF(bi) + dispIW(EK) + dispWW(RW)) -121.5, -132.1, -126.5, -138.2 kcal/mole, respectively, if the three and higher-body contribution is neglected, and (Table 7, SCF(SM) + dispIW(EK) + dispWW(RW)) -112.8, -114.7, -116.6, -121.1 kcal/mole when it is taken into account for the 4 configurations described above, the dispersion energy between the ion and the water molecules being similar for all these configurations from EK parameters. As commented above, it is difficult to state which is the most stable structure since the dispersion energy treatment is not corrected for the charge overlap effect in this region.

These results give a good idea of the limit of the validity of the conclusions which can be extracted from calculations using such approximations, as soon as unusual structures are considered.

The calculations have been performed on a IBM 3090 200VF machine, at the Centre de Calcul du CNRS de Strasbourg-Cronenbourg.

TABLE 7

 $\text{H}_3\text{O}^+(\text{H}_2\text{O})_6$: Total energy ^a

	3+3 PL	4+2 PL	3+3 PYR	4+2 PYR
MCY + EK	-124.7 ^b	-126.9	-140.5	-140.3
BB + AFHF	-91.5 ^b	-89.3	-98.6	-94.9
GGB + SPC	-94.8 ^b	-93.1	-99.2	-90.4
	-105.7 ^c	-102.5	-110.7	-101.0
SCF(bi) + dispIW(EK) + dispWW(SPC)	-115.5 ^b	-118.1	-116.9	-117.9
SCF(SM) + dispIW(EK) + dispWW(SPC)	-106.4 ^b	-100.7	-107.0	-100.9
SCF(bi) + dispIW(EK) + dispWW(RW) ^d	-121.5 ^b	-132.1	-126.5	-138.2
SCF(SM) + dispIW(EK) + dispWW(RW) ^d	-112.8 ^b	-114.7	-116.6	-121.1

a-b-c) : see footnotes on Table 3;

d) ref. 19

REFERENCES

- 1 J.D. Bernal, R.H. Fowler, J. Chem. Phys., **1** (1933) 515
- 2 H. Stillinger, A. Rahman, J. Chem. Phys., **60** (1974) 1545
- 3 H. Kistenmacher, G.C. Lie, H. Popkie, E. Clementi, J. Chem. Phys., **61** (1974) 546
- 4 O. Matsuoka, E. Clementi, M. Yoshimine, J. Chem. Phys. **64** (1976) 1351
- 5 H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, J. Hermans, in "Intermolecular Forces", edited by B. Pullman (Reidel, Dordrecht) 1981, p 331
- 6 E. Clementi, P. Habitz, J. Phys. Chem., **87** (1983) 2815
- 7 W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, J. Chem. Phys., **79** (1983) 926
- 8 I.P. Buffey, W. Byers Brown, Chem. Phys. Letters, **109** (1984) 59
- 9 E. Kochanski, Nouv. J. Chimie, **8** (1984) 605; Chem. Phys. Letters, **130** (1986) 291
- 10 Y. Guissani, B. Guillot, S. Bratos, J. Chem. Phys., **88** (1988) 5850
- 11 E. Kochanski, A. Rahmouni, J. Chim. Phys., **86** (1989) 995
- 12 E. Kochanski, Chem. Phys. Letters, **133** (1987) 143
- 13 N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, J. Chem. Phys., **21** (1953) 1087
- 14 S. Romano, E. Clementi, Gazz. Chim. Ital. **108** (1978) 319
- 15 G.H.F. Dierksen, W.P. Kraemer, B.O. Roos, Theor. Chim. Acta, **36** (1975) 249
- 16 E. Kochanski, E. Constantin, J. Chem. Phys., **87** (1987) 1661
- 17 S.F. Boys, F. Bernardi, Mol. Phys., **19** (1970) 553
- 18 W. Rijks, P.E.S. Wormer, J. Chem. Phys., submitted
- 19 A. Rahmouni, E. Kochanski, R. Wiest, P.E.S. Wormer, Chem. Phys. Letters, submitted
- 20 E. Kochanski, J. Amer. Chem. Soc., **107** (1985) 7869
- 21 Y.K. Lau, S. Ikuta, P. Kebarle, J. Amer. Chem. Soc., **104** (1982) 1462
- 22 M. Meot-Ner, C.V. Speller, J. Phys. Chem., **90** (1986) 6616
- 23 B.R. Gelin, M. Karplus, Biochemistry, **18** (1979) 1256; P.W. Weiner, P.A. Kollman, J. Comp. Chem., **2** (1981) 287; A. Vedani, J.D. Dunitz, J. Amer. Chem. Soc., **107** (1985) 7653; A. Vedani, J. Comp. Chem., **9** (1988) 269

DISCUSSION

Question - Is three-body contribution always repulsive ?

KOCHANSKI - Only the Self Consistent Field level three and higher-body contribution is considered here because it is much more important than the three-body dispersion energy. It is repulsive for these four geometrical configurations because the dominant water 1-ion-water 2 interactions are repulsive, water 1 and water 2 being two nearest water molecules of the first solvation shell. However, other terms may be attractive, in particular when one of the water molecules is in the second shell. They are generally smaller in magnitude. This means that an analytical description of the three-body contribution requires some repulsive AND attractive components.

MODELING OF ORGANOMETALLIC REACTIVITY USING A COMBINATION OF EXTENDED HÜCKEL AND MOLECULAR GRAPHICS TECHNIQUES

PIERRE-YVES MORGANTINI, PETER FLUEKIGER, JACQUES WEBER

Department of Physical Chemistry, University of Geneva, 30 quai Ernest Ansermet, 1211
Genève 4, Switzerland

AND E. PETER KÜNDIG

Department of Organic Chemistry, University of Geneva, 30 quai Ernest Ansermet, 1211
Geneva 4, Switzerland

SUMMARY

A new formalism has been developed in order to evaluate the intermolecular interaction energy E_{int} between an organometallic substrate S and an incoming reactant R in the framework of the extended Hückel method. Approximate procedures are used to estimate electrostatic, charge transfer and exchange repulsion components of E_{int} , the values of which are mapped as a local reactivity index onto the molecular surface of S by means of a standard color code. When applied to electrophilic and nucleophilic addition reactions such as the protonation of ferrocene or the attack of arene-Cr(CO)₃ by H⁺, this combination of quantum chemistry and molecular graphics techniques describes the high regioselectivity of these reaction mechanisms.

INTRODUCTION

Molecular graphics (MG), which can be defined as the application of computer graphics techniques to investigate molecular structure, function and interaction, is used today on a routine basis to build, represent and manipulate three-dimensional (3D) models of molecular structures and properties (ref. 1). As a rapid access to molecular geometry is a prerequisite, efficient procedures for structure and substructure searching within crystallographic data bases have been proposed (ref. 2). When no structural data are available, several powerful model builders based on molecular mechanics can be used in standard molecular modeling packages (ref. 3).

The calculation of molecular properties such as intermolecular interaction energies, however, usually requires time-consuming quantum chemical methods and faster alternatives must be derived for MG applications. For example, the well-known molecular electrostatic potential (MEP) (ref. 4) may now be approximated by using multipolar expansions of the electron density (ref. 5). This considerably reduces the computational effort required to map these MEP values onto molecular surfaces (ref. 6). Because of this, the MEP model has met with considerable success in modeling the complex situations arising from organic reaction mechanisms (ref. 7) or drug design applications (ref. 8). However, we have recently found (ref. 9) that the electrostatic component alone is not sufficient, even

qualitatively, for a proper description of reaction mechanisms involving an organometallic substrate S and an incoming reactant R with electrophilic or nucleophilic character. In such cases, there is generally an important charge transfer between the d shell of the transition metal atom of S and the orbitals of R and it is essential to take into account the corresponding energy which can be of the same order of magnitude as the electrostatic component. As the charge transfer mechanism is essentially due to orbital overlap, it is difficult to parametrize this energy component using an empirical potential. Instead, a quantum chemical calculation of the S-R supermolecule should be performed - which is considerably more demanding in terms of computer time, especially as this calculation should be repeated for all the possible locations of R on the molecular surface of S. Clearly, in view of the short response times required by MG applications, it was important to employ an efficient semi-empirical quantum chemical method to evaluate the charge transfer component. This is why we turned to the extended Hückel (EH) method (ref. 10) which is known to quickly and reasonably predict the electronic structure of transition metal complexes.

In this paper, after a brief description of the model we have developed, some recent results obtained for both electrophilic and nucleophilic additions to organometallic species will be reported and discussed.

METHOD

Within our model, the S-R interaction energy $E_{\text{int}}(\vec{r})$ is expressed as a sum of several components:

$$E_{\text{int}}(\vec{r}) = E_{\text{es}}(\vec{r}) + E_{\text{ct}}(\vec{r}) + E_{\text{ex}}(\vec{r}) \quad (1)$$

where \vec{r} specifies the position of the incoming electrophile or nucleophile reactant in the vicinity of the organometallic substrate; E_{es} , E_{ct} and E_{ex} being electrostatic, charge transfer and exchange energy components, respectively.

In order to display molecular surfaces color-coded according to E_{int} values used as a reactivity index, E_{int} is evaluated repeatedly at selected points \vec{r} located on the molecular surface of substrate S. The number of points depends, of course, on the size and complexity of the substrate, and in the cases presented here this number varies between 4'000 and 6'000. Negative (respectively positive) values of E_{int} correspond to S-R attractive (repulsive) interactions, and the regions where E_{int} is at a minimum are the most reactive sites of S to be attacked by R. In all cases, the color-coding range from red to yellow to blue extends smoothly over the numerical range of E_{int} from the most negative to zero to the most positive values, which means that the red zones correspond to preferred sites of attack. In this paper, however, the need for monochrome figures has lead us to suppress the shading of the surfaces and to map the red-yellow-blue color scale over the black-grey-white range.

It is important to choose a simple though realistic model for the reactant R, as the computer time required to evaluate E_{int} increases rapidly as a function of the complexity of R.

In order to have E_{int} values depending only on the position of R on the molecular surface of S, and not on its orientation, two spherically symmetric model reactants have been chosen: a proton with a virtual 1s orbital for the electrophile and an H^- hydride ion with two 1s electrons for the nucleophile. Let us turn now to the approximations we have used for the various components of E_{int} .

Electrostatic component

In the case of an electrophilic attack, E_{es} is equal to the MEP of substrate S:

$$E_{\text{es}}(\vec{r}) = \sum_A z_A / |\vec{r} - \vec{r}_A| - \sum_{\mu} \sum_{\nu} P_{\mu\nu} \langle \mu | 1/r | \nu \rangle \quad (2)$$

where the first term corresponds to nuclear repulsion, the summation running over all atoms A of S, with nuclear charge z_A located in \vec{r}_A . The second term originates from electronic attraction, $P_{\mu\nu}$ being the first order density matrix element corresponding to atomic orbitals (AOs) χ_{μ} and χ_{ν} , and $\langle \mu | 1/r | \nu \rangle$ being defined as

$$\langle \mu | 1/r | \nu \rangle = \int \chi_{\mu}(\vec{r}') \frac{1}{|\vec{r} - \vec{r}'|} \chi_{\nu}(\vec{r}') d\vec{r}' \quad (3)$$

In case of nucleophilic attack, we may reasonably assume that the electrostatic interaction between S and the H^- ion reduces to that between S and a negative point charge - which is obviously correct for rather large S-R distances where the so-called penetration integrals vanish.

Several test calculations have shown this to be the case for S-R distances roughly 1 Å larger than the usual van der Waals radii of the atoms of S, which explains why for nucleophilic attack we are using molecular surfaces of the substrates generated from larger atomic spheres (typically by 1 Å) than the standard van der Waals ones. On such surfaces, the electrostatic component for nucleophilic attack is therefore given by $-E_{\text{es}}$ (eqn. (2)).

As the calculation of $\langle \mu | 1/r | \nu \rangle$ integrals using the EH basis of atomic orbitals is time-consuming since they are of Slater type, we used the neglect of diatomic differential overlap (NDDO) approximation, according to which the second righthand term of eqn. (2) becomes:

$$\sum_{\mu} \sum_{\nu} P_{\mu\nu} \langle \mu | 1/r | \nu \rangle = \sum_A \sum_{\mu \in A} \sum_{\nu \in A} P_{\mu\nu} \langle \mu_A | 1/r | \nu_A \rangle \quad (4)$$

The first summation is over all atoms A of S. However, this requires the evaluation of the reduced density matrix using orthogonalized AOs.

Charge transfer component

It has recently been pointed out by Brown et al. (ref. 11) that organometallic substrates are often characterized by bands of closely spaced energy levels both in HOMO and LUMO regions. This renders the use of frontier orbitals only, in the evaluation of orbital or charge transfer effects, as questionable. Instead, Brown et al. have suggested replacing

the perturbational treatment by a complete EH calculation of the S-R supermolecule; the so-called "orbital interaction energy" is obtained as the difference between the total energies of the supermolecule and those of the separate fragments. We have used the same approach in our model.

As EH total energies fairly comprehensively represent the sum of the covalent energies within the chemical bonds, the S-R charge transfer energy may be approximated as:

$$E_{ct} = E^{tot}(S-R) - E^{tot}(S) - E^{tot}(R) \quad (5)$$

where $E^{tot}(X)$ represents the EH total energy of system X calculated as

$$E^{tot}(X) = \sum_i n_i \epsilon_i \quad (6)$$

with n_i and ϵ_i being occupation number and energy, respectively, of the i th MO of X.

$E_{ct}(\bar{r})$ is then readily obtained by positioning the reactant R at selected points \bar{r} on the molecular surface of S and by using expression (5).

Exchange component

In the case of electrophilic attack there is no exchange component in our model, as the reactant has no electrons; however, the situation is different for nucleophilic attack. For this case, the exchange term, which describes the short-range repulsion due to the overlap of both S and R electron distributions, is simply chosen to be zero outside the molecular surface of S and infinite on the surface itself (hard sphere approximation), which means that the minima of $E_{es} + E_{ct}$ on this surface are automatically taken as the most reactive sites. This approximation is actually justified by the $1/r^{12}$ behavior of the short-range repulsion component, which leads to a very steep function close to the nuclei.

To summarize, our model therefore rests on the following assumptions: polarization and dispersion energy components are neglected; the geometrical deformations of S when attacked by R are not taken into account; solvent effects are ignored so far. The model can thus be used to describe the initial stage of attack of mainly kinetically controlled processes exhibiting an early transition state.

Computational details

All the EH calculations have been performed using the single zeta Slater type atomic orbitals of Clementi and Roetti (ref. 12), except for the d shell of transition metal atoms which has been described by the double zeta functions of the same authors. For all hydrogen atoms and for the reactant, a 1s exponent of 1.0 has been used as it results in slightly better intermolecular interaction energies. For the evaluation of E_{es} , self-consistent charge and configuration (SCCC) calculations have been performed with a quadratic dependence of the valence state ionization energies (VSIE) of all the atoms. The calculation of E_{ct} , however, has been carried out without the SCCC procedure, with a VSIE of reactant H_{RR} systematically chosen at $\epsilon_{HOMO} + 0.2$ eV for the electrophilic attack and $\epsilon_{LUMO} - 0.5$ eV for

the nucleophilic attack, ϵ_{HOMO} and ϵ_{LUMO} being the HOMO and LUMO energies of the substrate, respectively. As H_{RR} has to be chosen $\epsilon_{\text{HOMO}} < H_{\text{RR}} < \epsilon_{\text{LUMO}}$, small deviations from the values optimized for soft reactants allow to take into partial account, when necessary, the very nature of R: hard electrophiles are characterized by larger (i.e. less negative) H_{RR} values and hard nucleophiles by smaller (i.e. more negative) H_{RR} values.

RESULTS AND DISCUSSION

Electrophilic addition to ferrocene

According to NMR studies (ref. 13), ferrocene protonates readily in strong acids to give the reaction intermediate $\text{Fe}(\text{C}_5\text{H}_5)_2\text{H}^+$. The protonation site is located on the metal, presumably in the equatorial plane. Ion thermochemistry experiments suggest that this is probably also the case in the gas phase, though ring protonation in an exo position is also possible with a proton affinity at least 5 kcal/mole smaller than on metal (ref. 14). A theoretical modeling of this reaction should, when solvent effects are not taken into consideration, predict the correct site of addition. The results obtained with our model are presented in Fig. 1. It is immediately seen that the five lowest minima of E_{int} , reflecting the

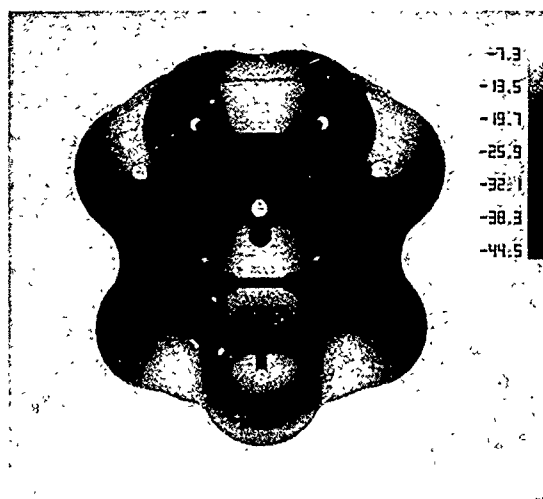


Fig. 1. Solid model of the molecular surface of ferrocene shaded according to the E_{int} property for electrophilic attack. The shading range from black to grey to white extends smoothly over the numerical range of E_{int} from the most negative to zero to the most positive values, which means that dark zones correspond to preferred sites of attack. The shading scale on the right indicates the E_{int} values mapping, in kcal/mol.

five-fold symmetry axis of the molecule, are located in equatorial positions of the molecular surface of $\text{Fe}(\text{C}_5\text{H}_5)_2$, corresponding therefore to protonation on metal. Furthermore, a detailed examination of Fig. 1 reveals that secondary minima, corresponding to less favorable interaction energies, are found on exo positions of the ligand rings, which

parallels the gas phase measurements. Interestingly, it is only by adding E_{ct} to E_{es} that the theoretical model correctly predicts the most stable protonation site. Indeed, the use of electrostatic effects alone as E_{int} leads to the ligand site predicted as the most favorable. As a similar conclusion has been drawn from several other examples we have been studying, the MEP model should be regarded with caution when used for predicting organometallic reaction mechanisms. This may be ascribed to the important charge transfer effects between substrate and reactant which are due to the d orbitals of the metal atom belonging to S.

Sequential nucleophilic and electrophilic addition to arene- $\text{Cr}(\text{CO})_3$

The structure and reactivity of arene-chromium tricarbonyl, $(\eta^6\text{-C}_6\text{H}_6)\text{Cr}(\text{CO})_3$, have been the subject of intensive investigations in organometallic chemistry (ref. 15). In this complex, as in most organometallic species containing unsaturated hydrocarbon ligands, metal-arene bonding leads to a net charge transfer from the ring to the metal, with the result that nucleophilic attack occurs readily on the exo-face of the positively charged hydrocarbon. One of us has recently shown that the $(\eta^6\text{-C}_6\text{H}_6)\text{Cr}(\text{CO})_3$ complex may undergo with high regioselectivity the following sequential reactions: (i) a nucleophilic addition by a reactive carbanion R^- to the exo-face of the benzene ring; (ii) an electrophilic addition to the intermediate anionic complex $(\eta^5\text{-C}_6\text{H}_6\text{R})\text{Cr}(\text{CO})_3^-$ directly to the metal atom; (iii) CO migratory insertion in the metal-electrophile bond, and (iv) reductive elimination to yield, after decomplexation, a 1,2-trans-disubstituted cyclohexadiene (ref. 16).

We have calculated the interaction energies between substrate and reactant for the first two steps of this sequential mechanism (ref. 17). Figure 2 shows that the initial nucleophilic

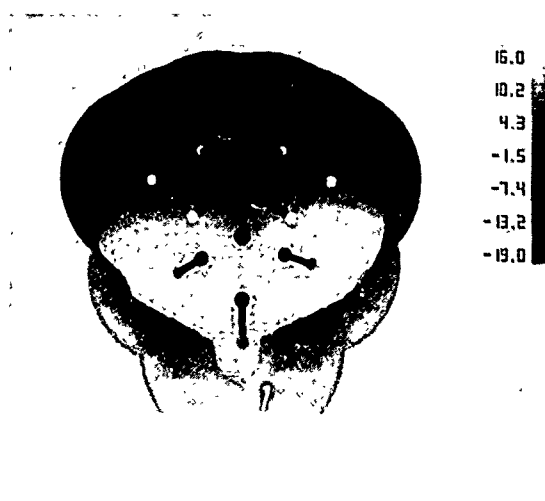


Fig. 2. Solid model of the molecular surface of arene- $\text{Cr}(\text{CO})_3$ shaded according to the E_{int} value for nucleophilic attack.

attack is likely to occur on the face opposite to the metal of a complexed arene, though the region with negative E_{int} values extends broadly over the upper part of the surface towards the metal atom.

The subsequent electrophilic attack has been modeled using an anionic cyclohexadienyl- $\text{Cr}(\text{CO})_3$ adduct with $\text{R} = 1,3\text{-dithian}$ (ref. 18). The results, presented in Fig. 3, indicate that the electrophilic attack should take place on the metal, which is

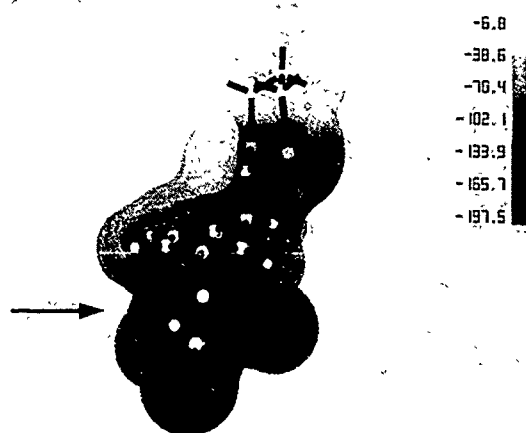


Fig. 3. Solid model of the molecular surface of the $\eta^5\text{-[6-(1,3-dithian-2-yl)cyclohexadienyl]tricarbonylchromium(0)}$ anion shaded according to the E_{int} value for electrophilic attack. The arrow indicates the position of the minimum of E_{int} .

in total agreement with a crystallographic study performed recently (ref. 18) on the $(\eta^5\text{-C}_6\text{H}_6\text{R})\text{Cr}(\text{CO})_3\text{R}'$ complex, with $\text{R} = 2\text{-methyl-1,3 dithian}$ and $\text{R}' = (\text{Ph})_3\text{Sn}$. In the crystal structure the metal- $\text{Sn}(\text{Ph})_3$ bond is trans to the cyclohexadienyl sp^3 carbon. Clearly, the present model, based on a combination of both extended Hückel and molecular graphics techniques, appears to be able to rationalize the important changes in reactivity which occur in organic fragments through metal complexation and, simultaneously, the high regioselectivity observed in sequential additions to such organometallic species. This, in turn, suggests that it can be used as a reliable tool in modeling the reactivity of such compounds.

ACKNOWLEDGEMENTS

The authors are grateful to Dr. C. Daul for fruitful discussions. In addition, they have benefited from the assistance of Mr. J. Leresche. This work is part of Project 20-25317.88 of the Swiss National Science Foundation.

REFERENCES

- 1 J.E. Dubois, D. Laurent and J. Weber, Chemical ideograms and molecular computer graphics, *Visual Computer*, 1 (1985) 49-63.
- 2 F.H. Allen, O. Kennard and R. Taylor, Systematic analysis of structural data as a research technique in organic chemistry, *Acc. Chem. Res.*, 16 (1983) 146-153.
- 3 U. Burkert and N. Allinger, *Molecular Mechanics*, American Chemical Society, Washington D.C., 1982.
- 4 E. Scrocco and J. Tomasi, The electrostatic molecular potential as a tool for the interpretation of molecular properties, *Top. Curr. Chem.*, 42 (1973) 95-170.
- 5 A. Goldblum, D. Perahia and A. Pullman, Use of the overlap multipole expansion for approximating molecular electrostatic potentials, *Int. J. Quant. Chem.*, 15 (1979) 121-129.
- 6 P.K. Weiner, R. Langridge, J.M. Blaney, R. Schaefer and P.A. Kollman, Electrostatic potential molecular surfaces, *Proc. Natl. Acad. Sci. USA*, 79 (1982) 3754-3758.
- 7 S.D. Kahn, C.F. Pau, A.R. Chamberlin and W.J. Hehre, Modeling chemical reactivity. 4. Regiochemistry and stereochemistry of electrophilic additions to allylic double bonds, *J. Am. Chem. Soc.*, 109 (1987) 650-663.
- 8 H.P. Weber, T. Lybrand, V. Singh and P.A. Kollman, Analysis of the pharmacological properties of clozapine analogues using molecular electrostatic potential surfaces, *J. Mol. Graphics*, 4 (1986) 56-60.
- 9 J. Weber, P. Fluekiger, P.Y. Morgantini, O. Schaad, A. Goursot and C. Daul, The modeling of nucleophilic and electrophilic additions to organometallic complexes using molecular graphics techniques, *J. Comp. Aid. Mol. Design*, 2 (1988) 235-253.
- 10 R. Hoffmann, An extended Hückel theory. I. Hydrocarbons, *J. Chem. Phys.*, 39 (1963) 1397-1412.
- 11 D.A. Brown, N.J. Fitzpatrick and M.A. McGinn, The prediction of nucleophilic attacking sites via determinants of active frontier and near-frontier orbitals, *J. Organomet. Chem.*, 293 (1985) 235-248.
- 12 E. Clementi and C. Roetti, Roothaan-Hartree-Fock atomic wavefunctions, *Atomic Data and Nuclear Data Tables*, 14 (1974) 177-478.
- 13 T.J. Curphey, J.O. Santer, M. Rosenblum and J.H. Richards, Protonation of metallocenes by strong acids. Structure of the cation, *J. Am. Chem. Soc.*, 82 (1960) 5249-5250.
- 14 M. Meot-Ner, Ion chemistry of ferrocene. Thermochemistry of ionization and protonation and solvent clustering. Slow and entropy driven proton-transfer kinetics, *J. Am. Chem. Soc.*, 111 (1989) 2830-2834.
- 15 A. Solladié-Cavallo, Arene-chromium tricarbonyl complexes: bonding and behaviour, *Polyhedron*, 4 (1985) 901-927.
- 16 E.P. Kündig, Recent advances in arene transformation reactions via chromium complexes, *Pure & Appl. Chem.*, 57 (1985) 1855-1864.
- 17 J. Weber, P.Y. Morgantini, J. Leresche and C. Daul, Molecular graphics as a tool for modelling nucleophilic and electrophilic reactions in organometallic chemistry, in: R. Carbo (Ed.), *Quantum Chemistry: Basic Aspects, Actual Trends*, Elsevier, Amsterdam, 1989, pp. 557-576.
- 18 E.P. Kündig, P. Paglia and G. Bernardinelli, manuscript in preparation.

DISCUSSION

PERAHIA - Avez-vous comparé vos résultats avec ceux obtenus en utilisant des méthodes ab initio ?

MORGANTINI - Oui, nous avons calculé pour plusieurs petites molécules (H_2O , NH_3 , CH_3F ...) l'énergie d'interaction pour des attaques électrophiles ou nucléophiles à l'aide de Gaussian 80 et de la décomposition d'énergie de Morokuma. Nous avons également fait des calculs similaires pour le ferrocène.

THE WATER DIMER - COMPARISON OF RESULTS OBTAINED BY BOTH AB INITIO SUPERMOLECULE AND SAPT METHODS. DERIVATION OF SIMPLIFIED FORMULAS

O. HESS¹, M. CAFFAREL¹, J. LANGLET¹, J. CAILLET¹, C. HUISZON²,
* P. CLAVERIE¹,

¹Dynamique des Interactions Moléculaires, Université Pierre et Marie Curie, 4 Place Jussieu, 75005 Paris (France)

²Chemical Physics Laboratory, Twente University of Technology, P. O. Box 217, 7500 AE Enschede, The Netherlands

+ Deceased, January 24, 1988.

SUMMARY.

Following explicit formulae for the calculation of second-order exchange contributions (induction as well as dispersion) within the framework of Symmetry-Adapted-Perturbation-Theories (SAPT), exchange contributions can be expressed as a combination of electrostatic interaction energies between suitably generalized charge distributions (overlap intermolecular distributions). Numerical calculations for the interaction of two water molecules are presented. The possibility of defining simple analytical functions representing each contribution of the interaction energy is discussed.

INTRODUCTION

In the field of theoretical evaluations of interaction energies, two types of approach are generally distinguished. The first approach is the so-called supermolecule method (ref. 1) in which the interaction energy is obtained as a difference between the total energy of the interacting molecules (the supermolecule) and the sum of the total energies of each monomer, all energies being calculated by using the same method. But the dispersion contribution cannot be obtained at the SCF level, in other words, an extensive CI calculation would be necessary to recover this important contribution. In the second approach, the intermolecular interaction energy is calculated from perturbation theory using the intermolecular potential as perturbing operator. When the intermolecular distance R is large, one is dealing with the Rayleigh-Schrödinger perturbation theory in which only simple products of monomer wavefunctions are used. Due to the large separation between monomers no antisymmetrization of the factorized wave functions is necessary. For shorter distances, e.g. distances corresponding to the region around the equilibrium configuration, the usual Rayleigh-Schrödinger perturbation theory must be abandoned (ref. 2) and in order to take into account, at least to some extent, the exchange of electrons between the interacting molecules, some form of exchange per-

turbation theory (the so-called Symmetry Adapted Perturbation Theories (SAPT), (refs. 3-4) must be used. It is important to emphasize that this approach is particularly attractive with regard to the usual supermolecular approach since the interaction energy is decomposed into a sum of terms for each of which it is possible to give some physical interpretation (at least for terms up to and including second-order terms). This is a very appealing feature for a qualitative understanding of the interaction and can be very helpful for the development of simplified formulas for intermolecular interactions.

To our knowledge, the first example of an exchange-perturbation theory calculation is due to Jeziorski and van Hemert in their pioneering work on the water dimer (ref. 5). Neglecting all intramonomer correlation effects, they evaluated the complete first-order interaction energy $E^{(1)} = E_{RS}^{(1)} + E_{exch}^{(1)}$ (explicitly, the sum of the Rayleigh-Schrödinger and first-order exchange energies) and only the Rayleigh-Schrödinger second-order interaction energy.

Very recently Hess *et al.* (ref. 6) have presented a new method of deriving explicit formulas for the calculation of second-order exchange contributions (induction as well as dispersion) within the framework of Symmetry-Adapted-Perturbation-Theory. Numerical results for the interaction of two water molecules have been presented, putting into evidence the nonnegligible role of the complete second-order exchange contributions. But it is well known that the quality of the results strongly depends on the size of the basis set used in the calculations, thus such a method cannot be applied to arbitrarily large systems. In fact the ability to determine with a high accuracy the values of each component of intermolecular interaction energy opens a way towards representing them through simple analytical functions fitted on values calculated in the framework of this perturbation treatment. In this present work we have been interested by the development of simplified formulae for the calculation of the dispersion and exchange-dispersion energies.

The organization of the present paper is as follows. In Sec. I and II we summarize the formal development of the second-order exchange contributions derived by Hess *et al.* (ref. 6) together with the most important results obtained by these authors for the water dimer. Sec. III is devoted to the investigation of basis set effects upon the different interaction energy components calculated with the method here-above cited. In section IV we will present and discuss some simplified formulae elaborated for dispersion and exchange dispersion contributions.

I. METHOD

We will just summarize the formal development of second-order exchange contributions presented by Hess *et al.* (ref. 6)

Following standard Symmetry-Adapted Perturbation Theories (refs. 3,4) the complete first- and second-order interaction energies are written as:

$$E^{(1)} = \frac{\langle \Psi_0^A \Psi_0^B | V^{AB} A | \Psi_0^A \Psi_0^B \rangle}{\langle \Psi_0^A \Psi_0^B | A | \Psi_0^A \Psi_0^B \rangle}, \quad (1)$$

$$E^{(2)} = - \frac{\langle \Psi_0^A \Psi_0^B | V^{AB} R_0 A (V^{AB} - E^{(1)}) | \Psi_0^A \Psi_0^B \rangle}{\langle \Psi_0^A \Psi_0^B | A | \Psi_0^A \Psi_0^B \rangle}, \quad (2)$$

where R_0 denotes the reduced resolvent of H_0 given by

$$R_0 = \sum'_{ij} \frac{|\Psi_i^A \Psi_j^B\rangle \langle \Psi_i^A \Psi_j^B|}{(E_i^A + E_j^B) - (E_0^A + E_0^B)}, \quad (3)$$

(the prime in \sum' means as usual that the term corresponding to $i=0$ and $j=0$ is excluded from the summation) and A is the intersystem antisymmetrizer which is written in the form (ref. 7).

$$A = 1 - A' = 1 - P_{(1)} + P_{(2)} - \dots + (-1)^{N_{ex}} P_{(N_{ex})}, \quad (4)$$

where $P_{(n)} = \sum_i \sum_j P_{ij}$ denotes the sum of all permutations exchanging (space and spin) coordinates of electron i of molecule A with coordinates of electron j of molecule B, and similar definitions hold for $P_{(2)}, P_{(3)}, \dots$ (N_{ex} denotes the smallest value of N_A and N_B , the numbers of electrons of molecule A and B respectively).

The second-order perturbation energy $E^{(2)}$ (Eq. (2)) may be decomposed into the usual second-order Rayleigh-Schrödinger (RS) perturbation energy $E_{RS}^{(2)}$ (obtained by putting $A = 1$ in Eq. (1)) and into the so-called second-order exchange energy $E_{exch}^{(2)}$

$$\begin{aligned} E_{exch}^{(2)} &\equiv E^{(2)} - E_{RS}^{(2)} \\ &= - \frac{\langle \Psi_0^A \Psi_0^B | (V^{AB} - E^{(1)}) (A' - \langle A' \rangle) | \Phi^{(1)} \rangle}{\langle A \rangle}, \end{aligned} \quad (5)$$

where $\langle A' \rangle$ and $\langle A \rangle$ are the expectation values of A' and A calculated with the ground-state wavefunction $\Psi_0^A \Psi_0^B$ and $\Phi^{(1)}$ stands for the first-order correction to the wavefunction in the perturbation theory (ref. 4).

$$\Phi^{(1)} = -R_0 V^{AB} \Psi_0^A \Psi_0^B. \quad (6)$$

Now, since multiple exchanges are supposed to contribute weakly in the region around the equilibrium geometry (ref. 8-9) only the leading contribution to $E_{exch}^{(2)}$ corresponding to a single exchange of electrons between molecules A and B has been calculated. Thus, putting $A' = P_{(1)}$ in Eq. (5) and neglecting terms which will correspond to contributions of order higher than S^2 (where S stands for overlap integrals between orbitals of monomers A and B) within the Hartree-Fock formalism used below, $E_{exch}^{(2)}$ is obtained

$$E_{exch}^{(2)} = - \langle \Psi_0^A \Psi_0^B | (V^{AB} - \langle V^{AB} \rangle) (P_{(1)} - \langle P_{(1)} \rangle) | \Phi^{(1)} \rangle, \quad (7)$$

Rewriting $\Phi^{(0)}$ (Eq. (6)) as follows

$$\Phi^{(1)} = \Psi_0^A \Phi_{ind}^B + \Phi_{ind}^A \Psi_0^B + \Phi_{disp}^{AB} \quad (8)$$

and inserting the previous decomposition of $\Phi^{(0)}$ into Eq. (7), the second-order exchange energy decomposes into three terms

$$E_{exch}^{(2)} = E_{exch-ind}^{(2)}(A \rightarrow B) + E_{exch-ind}^{(2)}(B \rightarrow A) + E_{exch-disp}^{(2)} \quad (9)$$

The sum of the first two terms in Eq. (9) will be referred to in the following as the exchange induction energy, while $E_{exch-disp}^{(2)}$ will be referred to as the exchange dispersion energy.

In fact following Claverie (ref. 7) the method adopted by Hess *et al.* (ref. 6) is essentially to express exchange contributions as a combination of formal electrostatic interaction energies between suitably generalized charge distributions (so called overlap intermolecular charge distributions). To do that, two basic ingredients have been used, namely:

1) The possibility of reducing the action of intersystem antisymmetrizer (appearing in SAPT) on factorized SCF wave functions to a sum of simple products of SCF determinants pertaining to each subsystem, namely

$$P_{(1)}[\Psi^A \Psi^B] = \sum_{i \in A} \sum_{j \in B} \Psi^A \begin{pmatrix} b_j \\ a_i \end{pmatrix} \Psi^B \begin{pmatrix} a_i \\ b_j \end{pmatrix} \quad (10)$$

where the summation is over the spin-orbitals of determinants Ψ^A (here labeled by i) and Ψ^B (labeled by j). Using Eq. (10) all integrals involving functions of the type $P_{(1)}[\Psi^A \Psi^B]$ are reduced to sums of integrals involving simple products $\Psi^A \begin{pmatrix} b_j \\ a_i \end{pmatrix} \Psi^B \begin{pmatrix} a_i \\ b_j \end{pmatrix}$ of "opposite transfer" determinants.

2) The next step consists of the use of the so-called Longuet-Higgins representation of the interaction operator V^{AB} in terms of the molecular charge distributions ρ^M ($M = A, B$) (ref. 10), namely:

$$V^{AB} = \iint \frac{\rho^A(\vec{r}^A) \rho^B(\vec{r}^B)}{|\vec{r}^A - \vec{r}^B|} d\vec{r}^A d\vec{r}^B \quad (11)$$

with

$$\begin{aligned} \rho^M(\vec{r}) &= \rho_{nuclear}^M(\vec{r}) + \rho_{electronic}^M(\vec{r}) \\ &= \sum_{\mu \in M} Z_{\mu} \delta(\vec{r} - \vec{r}_{\mu}) - \sum_{i \in M} \delta(\vec{r} - \vec{r}_i) \quad M = A, B \end{aligned} \quad (12)$$

The use of these two ingredients allows this development to have a systematic character. Within the SCF approximation, the different components of interaction

energy are now written as appropriate combinations of (mono- bielectronic and overlap) integrals involving spin-orbitals of closed-shell isolated systems (for more details see ref. 6).

Now, when performing the practical evaluation of the quantities necessary to the above calculation, one is faced to the well-known problem of summing expressions defined over the infinite set of unoccupied orbitals of the Fock operator belonging to the continuous spectrum. As pointed out by Jeziorski and van Hemert (ref. 5), such summations are practically inexecutable integration. To overcome this difficulty, it has been used the variational-perturbation method proposed by these authors. This method which is essentially based on the minimization of a Hylleraas-type functional, has been already described in detail (see e.g. refs. 5,11)

II. NUMERICAL RESULTS AND DISCUSSION

All calculations have been done for a fixed relative orientation of the two interacting water molecules and by varying only the distance R_{OO} between the two oxygen atoms. In order to facilitate comparisons, the fixed orientation has been chosen to be identical with that used by Jeziorski and van Hemert in their original work on the water dimer (ref. 5). Calculations have been performed using a substantially larger basis set. The so-called isotropic part of the basis (functions describing orbitals occupied in the ground-states of the atoms, see ref. 12) has been taken from ref. 13 and consists of a set of (13s8p) and (6s) functions on the oxygens and hydrogens, respectively. This basis set has been extended with a set of (2d) and (2p) polarization functions on oxygen and hydrogen respectively. The exponents were chosen in order to minimize the dispersion as well as the complementary exchange energies (see ref. 7). Exponents $\alpha_d = 1$ and 0.3 , $\alpha_p = 0.6$ and 0.15 have been obtained. The complete contracted basis consists of 94 basis functions for the water dimer.

The energy of the water monomer calculated by using this basis set equals -76.06004 a.u.. The SCF binding energies obtained for the water dimer are -3.96 kcal/mol and -3.73 kcal/mol without and with the counterpoise correction (CP), respectively. The latter value agrees very well with the SCF limit of -3.73 ± 0.05 kcal/mol (including CP correction) recently estimated by Szalewicz *et al* (ref. 14) using a very large basis set containing 212 contracted orbitals. The values of the particular contributions to the interaction energy are listed in Table I.

The essential results to point out are the following ones:

- 1) the second order exchange-induction was found to be quite important. At equilibrium distance, it compensates approximatively for 50% the induction energy. The importance of this contribution has been already noticed for inert gas dimers (refs. 9,11,15,16).

2) the second order exchange-dispersion energy represents about 20% of the dispersion energy, thus confirming the non-negligible role of this contribution.

R_{oo}^b	$E_{RS}^{(1)}$	$E_{exch}^{(1)}$	$E_{ind}^{(2)}$	$E_{disp}^{(2)}$	$E_{exch-ind}^{(2)}$	$E_{exch-disp}^{(2)}$
4.40	-23.66	50.31	-21.25	-8.90	14.28	3.32
4.80	-16.68	24.19	-9.42	-5.27	6.19	1.51
5.20	-10.81	11.61	-4.37	-3.18	2.70	0.75
5.67	-6.89	4.85	-1.82	-1.79	0.99	0.32
7.00	-2.67	0.39	-0.22	-0.46	0.06	0.03
9.00	-1.05	0.01	-0.03	-0.09	0.00	0.00

Table I. Particular contributions to the interaction energy of the water dimer (in kcal/mol) calculated with a 94 AO basis set.^a

a. Basis set described in the text.

b. Atomic units.

It has seemed interesting to compare the SCF binding energy to the sum of the complete first-order and second-order induction energies; these values are displayed in table II (columns 3 and 2 respectively).

R_{oo}^a	$E_{RS}^{(1)} + E_{ind}^{(2)} + E_{exch-ind}^{(2)}$	E_{int}^{SCF}	$E_{int}^{pert}^b$	$E_{int}^{SCF} + E_{disp}^{(2)}$	$E_{int}^{SCF} + E_{disp}^{(2)} + E_{exch-disp}^{(2)}$
4.40	16.68	10.97	11.10	2.07	5.39
4.80	4.28	1.08	0.52	-4.19	-2.68
5.20	-0.87	-2.62	-3.30	-5.80	-5.05
5.67	-2.87	-3.73	-4.34	-5.52	-5.20
7.00	-2.44	-2.55	-2.87	-3.01	-2.98
9.00	-1.07	-1.08	-1.15	-1.16	-1.16

Table II. Comparison of the SCF and perturbation theory interaction energies for the water dimer (in kcal/mol) calculated with the 94 AO basis set.

a. Atomic units.

b. Pure perturbational interaction energy calculated as :

$$E_{int}^{pert} = E_{RS}^{(1)} + E_{exch}^{(1)} + E_{ind}^{(2)} + E_{disp}^{(2)} + E_{exch-ind}^{(2)} + E_{exch-disp}^{(2)}$$

Except at large distances, Table II clearly demonstrates the non-coincidence of these two quantities. In fact the additional terms present in the SCF binding energy (induction part of third and higher-order Rayleigh-Schrödinger terms, some intramolecular correlation contribution introduced when doing a SCF supermolecule calculation

(ref. 8)) contribute in a non negligible way, even in the neighborhood of the equilibrium geometry. It may be noticed that these additional contributions become more and more important as the intermolecular distance is decreased. One might expect that the difference between E_{int}^{SCF} and $E^{(0)} + E_{ind}^{(2)}$ would be partly cancelled if in the perturbational approach the induction part of third and higher-order contributions would be considered. Thus presently, as a possibility, one may calculate the interaction energy by adding to the SCF binding energy the dispersion term calculated within a perturbation method, but in that case one has to be cautious to also take into account the exchange-dispersion terms. Thus one has to use the following decomposition:

$$E_{int} \approx E_{int}^{SCF} + E_{disp}^{(2)} + E_{exch-disp}^{(2)} \quad (13)$$

Table II shows that the location of the energy minimum is different following that the second order exchange-dispersion term is (or is not) taken into account. The values we have obtained are $R_{0.0} = 5.67$ a.u. and 5.20 a.u. respectively.

III. IMPORTANCE OF THE BASIS SET QUALITY

Now, we will pay some attention to the important problem of the quality of the basis set used. The results listed in table III clearly show that the different components of perturbation development are quite sensible to the choice of the basis set.

Base	STO-3G	4-31G	6-31G**	JvII ^a	O.II ^b
$E_{RS}^{(0)}$	-4.12	-8.88	-7.13	-7.11	-6.89
$E_{exch}^{(0)}$	1.72	2.10	2.60	4.89	4.85
$E_{ind}^{(2)}$	-0.64	-0.94	-1.04	-1.63	-1.82
$E_{disp}^{(2)}$	-0.20	-0.42	-0.93	-1.54	-1.79
$E_{ind-exch}^{(2)}$	0.45	0.25	0.33	0.80	0.99
$E_{disp-exch}^{(2)}$	0.04	0.05	0.13	0.27	0.32
E^{total}	-2.75	-7.84	-6.04	-4.32	-4.34
Dipole moment	1.72	2.60	2.18	2.06	1.98

Table III. Different components of intermolecular interaction energies calculated within different basis sets. All energy values are given in kcal/mol. The dipole moments are given in Debyes. (a) JvII stands for a gaussian basis (11,7,2/6,1) contracted into (4,3,2/2,1), this means 70 functions for the dimer (ref. 5). (b) This basis set containing 94 AO for the dimer has been discussed in section III. Results obtained using the geometry of ref. 5 with $R_{0.0} = 5.67$ a.u.

1) The electrostatic component is well reproduced only if the wave function of the unperturbed system correctly describes the charge distribution of isolated molecules (monomers). The calculation of multipolar moments gives a good criterion for the quality of the basis set. In this work we have calculated the dipole moment of the water molecule within the different basis sets studied, results presented in Table 3 (last line) have to be compared to the experimental value of 1.85 Debyes (ref. 17). When using the very large basis set (47 AO for water monomer), we have obtained a value which is in a very good agreement with the Hartree-Fock limit value (1.98 Debyes) estimated by Szalewicz et al. (ref.14). In fact, it is well known that at the SCF level, dipole moments are calculated with an error of about 10% because of the lack of electronic correlation at this level (ref. 12,14).

2) The exchange energy increases with the size of the basis set used; even 6-31G** basis set underestimates this contribution. This result proceeds from the imperfect behaviour of the wave function at long range leading to an underestimation of overlap effects between electronic clouds of different sub-systems. It has been noticed that nearly same values have been obtained when using basis sets including 70 and 94 AO for a water dimer.

It has appeared interesting to compare the first order energy we have calculated with the so-called Heitler-London first order energy defined as:

$$E_{HL}^{(1)} = \frac{\langle \Psi_0 | H | \Psi_0 \rangle}{\langle \Psi_0 | \Psi_0 \rangle} \quad (14)$$

where Ψ_0 is the antisymmetrized wave function calculated with the exact wave function of the two monomers,

$$\Psi_0 = A \Phi_0 = A \Psi_0^A \Psi_0^B \quad (15)$$

and H the total Hamiltonian of the system in interaction. In (in and only) that case the Heitler London energy may be written as:

$$E_{HL}^{(1)} = E_0 + \frac{\langle \Phi_0 | V^{AB} | A \Phi_0 \rangle}{\langle \Phi_0 | A | \Phi_0 \rangle} \quad (16)$$

where E_0 is the eigenvalue of H_0 , and the second term of this equation is nothing else that the total first order perturbation component. In practice we do not use exact wave functions for calculating isolated molecules, so Eq.16 is written as:

$$E_{HL}^{(1)} = \bar{E}_0 + \frac{\langle \Phi_0 | (\bar{E}_0 - H_0) A' | \Phi_0 \rangle}{\langle \Phi_0 | A | \Phi_0 \rangle} + \frac{\langle \Phi_0 | V^{AB} A | \Phi_0 \rangle}{\langle \Phi_0 | A | \Phi_0 \rangle} \quad (17)$$

where \bar{E}_0 represents the mean energy associated to approximate wave functions of the monomers. The last term of the second member of Eq.17 is the first order perturbation contribution including both Rayleigh-Schrodinger and exchange terms, the second term

(second member) called *complementary exchange* energy represents a correction which is zero when Ψ_0 is the exact eigen function of H_0 . Now, the Heitler-London energy is now decomposed into:

$$E_{HL}^{(1)} = \bar{E}_0 + E_{RS}^{(1)} + E_{exch}^{(1)} + E_{exch-compl}^{(1)} \quad (18)$$

We may denote:

$$E_{exch-HL}^{(1)} = E_{exch}^{(1)} + E_{exch-compl}^{(1)} \quad (19)$$

The value of the complementary exchange energy is a good mean to check the quality of the basis set. (For more details see ref. 7, 17). Table IV displays the calculated values of $E_{exch}^{(1)}$ and $E_{exch-compl}^{(1)}$. It appears that only very large basis sets (at least 70 AO for a water dimer) lead to quite correct values of first order exchange contributions.

Basis	STO-3G	4-31G	6-31G**	JvII ^a	O.II ^b
$E_{exch}^{(1)}$	1.72	2.10	2.60	4.89	4.85
$E_{exch-compl}^{(1)}$	1.74	1.48	1.12	-0.32	-0.17
$E_{exch-HL}^{(1)}$	3.46	3.58	3.72	4.57	4.68

Table IV. Dependence of the contributions $E_{exch}^{(1)}$, $E_{exch-compl}^{(1)}$, $E_{exch-HL}^{(1)}$ with the basis set. For (a) and (b) see comments in Table III. All values (in kcal/mol) have been calculated for the geometry of ref. 5 and for $R_{O-O} = 5.67$ a.u.

3) The total induction and dispersion terms are correctly taken into account only if very large basis sets are used. We may notice that the basis consisting in 94 AO which includes neither f-orbitals on oxygen nor d-orbitals on hydrogen leads to a value of the dispersion term which is very close to the limit value estimated by Szalewicz et al. (-2.kcal/mol) (ref. 14)

IV. DERIVATION OF SEMI-EMPIRICAL FORMULAE FOR SECOND ORDER DISPERSION CONTRIBUTION

These two contributions have been calculated as a sum of atom-atom interactions :

$$E_X^{(2)} = \sum_{i \in A} \sum_{j \in B} E_X(i,j) \quad (20)$$

where i (and j) are atoms belonging to molecule A (and B); the subscript X stands for dispersion or exchange-dispersion.

1) Second-order dispersion energy

This contribution is represented by:

$$E_{disp}^{(2)}(i,j) = - \left(\frac{C_6}{z^6} + \frac{C_8}{z^8} + \frac{C_{10}}{z^{10}} \right) k_i k_j \quad (21)$$

with $z = R_y/R_y^0$ and $R_y^0 = [(2R_i^*)(2R_j^*)]^{1/2}$ where R_i^* and R_j^* are the van der Waals radii of atom i and j . R is the distance between atoms i and j . Factors k_i, k_j allow the energy minimum of $E = E_{disp} + E_{exch}$, to have different values according to the atomic species involved (ref. 21).

Coefficients C_6, C_8 , and C_{10} have been calculated by identifying Eq. (21) with the one given by Stogryn *et al.* (ref. 19) for He...He interaction:

$$E_{disp}^{(2)} = - \left[\frac{1.471}{R^6} + \frac{14.1}{R^8} + \frac{182.0}{R^{10}} \right] \quad (22)$$

Namely: $C_6 = 0.143 \text{ kcal } \text{\AA}^6/\text{mol}$, $C_8 = 0.0381 \text{ kcal } \text{\AA}^8/\text{mol}$, $C_{10} = 0.0137 \text{ kcal } \text{\AA}^{10}/\text{mol}$.

The terms besides C_6/R^6 are not negligible in the equilibrium distance. For He...He, for instance, these two terms amount to 1/3 of the main term $-C_6/R^6$.

But it is well known that the multipolar part ($-1/R^n$) of the dispersion energy overestimates it at short distances owing to the neglect of the penetration part of the intermolecular integrals which appear in the numerators of the perturbation expansion (for more details see ref. 20 and references therein). In order to take into account the reduction of different multipolar terms, we have applied the process defined by Caillet *et al.* (ref. 21) when dealing with the sixth order power term of dispersion. Namely, we choose two distances, $R_M = R_i^* + R_j^*$ and $R_m = \frac{1}{2.42} R_M$. Then for $R > R_M$ we use the normal parameters C_n ($n=6, 8, 10$), for $R < R_m$ we use modified reduced parameters C'_n and for $R_m < R < R_M$ we use interpolated values of these parameters according to:

$$C_n(x) = \frac{(C_n + C'_n)}{2} + (0.375x^5 - 1.25x^3 + 1.875x) \frac{(C_n - C'_n)}{2} \quad (23)$$

where:

$$x = \left[R - \frac{(R_M + R_m)}{2} \right] / \left[\frac{(R_M - R_m)}{2} \right]$$

The polynomial $P(x)$ has been chosen in order that: a) $P(1) = 1$ and $P(-1) = -1$;

b) the first and second derivatives of $P(x)$ are continuous.

$C'_6 = C_6/6.25$; $C'_8 = C_8/7.38$ and $C'_{10} = C_{10}/10.44$.

Using the geometrical arrangement studied by Jeziorski *et al.* (ref. 5) and varying only R_{O-O} , $E_{disp}^{(2)}$ has been fitted with regards to values calculated by SAPT method.

2) Second-order exchange energy

It has been found that this component which is purely short-range varies exponentially with the distance R . The best fit has been given by the following analytical function

$$E_{exch-disp}^{(2)} = k_i k_j \left(1 - \frac{Q_i}{N_i^{val}}\right) \left(1 - \frac{Q_j}{N_j^{val}}\right) C e^{-\alpha z} \quad (24)$$

where Q_x ($x = i$ or j) is the net charge of atom x and N_x^{val} the number of valence electrons of atom x .

In the same way, as recommended by Caillet *et al.* (ref. 21), we have used the factor $(-Q_i)/N_i^{val}$ corresponding to the influence of the real electronic population of each atom on short range terms.

$C = 484.98$ kcal/mol and $\alpha = 9.18$.

3) Results.

R_{00}	$E_{disp-M}^{(2)}$	$E_{disp}^{(2)}$	$E_{disp-exch}^{(2)}$
4.50	-20.87	-7.65 (-7.80)	2.57 (2.92)
4.80	-13.88	-5.04 (-5.27)	1.52 (1.51)
5.00	-8.73	-3.98 (-4.10)	1.07 (1.07)
5.20	-5.77	-3.20 (-3.18)	0.75 (0.75)
5.67	-2.34	-1.91 (-1.80)	0.33 (0.32)
6.00	-1.47	-1.33 (-1.22)	0.19 (0.18)
7.00	-0.42	-0.42 (-0.47)	0.03 (0.03)

Table V. Values (in kcal/mol) calculated with simplified formulas, and by ab initio SAPT method (in parenthesis). For notations see the text.

In Table V, we have listed the values of the multipolar part of the second order dispersion denoted $E_{disp-M}^{(2)}$ (Eq. 21), the values obtained for $E_{disp}^{(2)}$ when the penetration part is taken into account and the value of second order exchange-dispersion terms $E_{disp-exch}^{(2)}$, the values between brackets are the ones calculated, within SAPT method (6). It may be noticed that even in the equilibrium region the taking in account of only the multipolar part of dispersion overestimates this contribution. The agreement between values calculated using simplified formulae or SAPT method is quite good in the case of second order exchange-dispersion energy and good enough as concerns the second order dispersion term.

IV. CONCLUSION

At this point, it is important to emphasize that the goal of this work was not to obtain a very accurate value of the interaction energy between two water molecules, since it is clear that for such a simple system, the supermolecule approach based on very large CI (refs. 10,11) are preferable. Actually, one of the basic motivation of our work was to demonstrate the nonnegligible role of the complete second-order exchange contribution (exchange induction as well as exchange-dispersion components). Furthermore the ability to determine quantitatively the importance of each component of the total interaction energy has opened the way towards representing them through simple analytical functions fitted on calculated values. It has appeared that such a fitting has to be done departing from results obtained within a very large basis set. But as it has been discussed in this paper the induction part of third (and perhaps higher orders) contributions should be considered. In fact, these contributions may be obtained from SCF results. As a first step simplified formulas for the calculation of the second-order dispersion (including exchange-dispersion) terms have been given. Work is under progress in order to verify other contributions (mainly first order exchange energy) of interaction energy. Such a possibility is essential with respect to the problem of elaborating *high quality simplified* functions for the calculation of the interaction energy between *arbitrarily large* molecules.

ACKNOWLEDGEMENT

The authors wish to thank the Groupement Scientifique "Modélisation Moléculaire" IBM-CNRS for providing them with computer facilities on IBM 3090/600E.

REFERENCES

1. K. Morokuma and K. Kitaura, Variational approach (SCF ab initio calculations) to the study of Molecular Interactions, in *Molecular Interactions*, Variational approach (SCF ab initio calculations) to the study of molecular interaction. edited by H. Ratajczak and W.J. Orville-Thomas (Wiley, New-York, 1980), Vol. 1, pp. 21-65.
2. P. Claverie, Theory of Intermolecular Forces. I-On the inadequacy of the usual Rayleigh-Schrödinger Perturbation Method for the Treatment of Intermolecular Forces, *Int. J. Quantum Chem.* 5 (1971) 273-295.
3. B. Jeziorski and W. Kolos, Perturbation approach of weak intermolecular interactions, in *Molecular Interactions*, edited by H. Ratajczak and W.J. Orville-Thomas (Wiley, New-York, 1982), Vol. 3, pp. 1-46.
4. I.G. Kaplan in *Theory of Molecular Interactions*, studies in Physical and Theoretical Chemistry, (Elsevier, Amsterdam, 1986), Vol. 42.
5. B. Jeziorski and M. van Hemert, Variation-perturbation treatment of the hydrogen bond between two water molecules, *Mol. Phys.* 31 (1976) 713-729.

6. O. Hess, M. Caffarel, C. Huiszoon and P. Claverie, Second Order Exchange Effect in Intermolecular Interactions. The Water Dimer, accepted for publication by J. Chem. Phys.
7. P. Claverie in *Intermolecular Interactions: From Diatomics to Biopolymers*, edited by B. Pullman (Wiley, New-York, 1978).
8. G. Chalasinski and B. Jeziorski, Exact calculation of Exchange Polarization energy for H_2^+ , Int. J. Quantum Chem. 7 (1973) 63-73.
9. G. Chalasinski and B. Jeziorski, On the exchange polarization effects in the interaction of two helium atoms, Mol. Phys. 32 (1976) 81-91.
10. H.C. Longuet-Higgins, The electronic states of composite systems, Proc. Roy. Soc. 235 (1956) 537-543.
11. G. Chalasinski, Perturbation calculations of the interaction energy between closed-shell Hartree-Fock atoms-The Neon dimer Mol. Phys. 49 (1983) 1353-1373.
12. J.G.C.M. van Duijneveldt-van de Rijdt and F.B. van Duijneveldt, Gaussian basis sets which yield accurate Hartree-Fock electric moments and polarizabilities. J. Mol. Struct. Theochem. 89 (1982) 185-201.
13. R. Poirier, R. Kari, and I.G. Csizmadia *Handbook of gaussian basis sets*, Physical sciences data 24 (Elsevier, 1985) Tables 8.76.2 and 1.41.2
14. K. Szalewicz, S.J. Cole, W. Kolos and R.J. Bartlett, A theoretical study of the water dimer interaction, J. Chem. Phys. 89 (1988) 3662-3673.
15. G. Chalasinski and B. Jeziorski, The Beryllium-Beryllium interaction exchange Polarization Effects in the interaction of closed shell systems, Theoret. Chim. Acta 46 (1977) 277-290.
16. G. Chalasinski, Exchange-Perturbation Calculation of the Interaction energy between Be atoms including intra-atomic correlation effects, Chem. Phys. 82 (1983) 207-213.
17. S.A. Clough, Y. Beers, G.P. Klein, and L.S. Rothman, Dipole moment of water from Stark measurements, J. Chem. Phys. 59 (1973) 2254-2259.
18. O. Hess, Approche perturbative des interactions moléculaires. Calcul ab initio et Monte-Carlo quantique des composantes de l'énergie d'interaction. Thesis, Université Paris VI, 1989.
19. D.E. Stogryn, High order interaction energies for systems for asymmetric molecules, Mol. Phys. 22 (1971) 81-103.
20. P. Claverie, Molecular Charge Distributions and Response Functions: Multipolar and Penetration Terms; Application to the theory of Intermolecular Interactions in Molecules in Physics, Chemistry and Biology, by Kluwer Academic Publishers (1988) Vol II pp. 393-415.
21. J. Caillet, P. Claverie and B. Pullman, On the conformational varieties of Adrenaline. The free molecule and the molecule in the crystal, Acta Cryst. B32 (1976) 2740-2745.

**WATER MOLECULE IN DIFFERENT CRYSTAL SURROUNDINGS:
PERIODIC HARTREE-FOCK AND MODEL HAMILTONIAN CALCULATIONS**

J. G. ÁNGYÁN and B. SILVI

*Laboratoire de Spectrochimie Moléculaire. (UA 508), Université Pierre et Marie Curie,
4, Place Jussieu, 75230 PARIS Cedex. (France)*

ABSTRACT. The effect of surroundings in molecular crystals has been studied at three levels of approximation: by periodic Hartree-Fock method, by self-consistent Madelung potential approach and by the usual cluster (prototype molecule) model. The calculations performed on different proton-ordered ice modifications and on lithium hydroxide monohydrate show that the SCMP method provides an electron distribution in very good agreement with the periodic Hartree-Fock results, while finite size effects induce considerable discrepancies within the cluster approach.

1. Introduction

It is a well-established fact that intermolecular forces may have sometimes considerable effects on the properties of individual molecules. It is especially true, when one is dealing with molecules in a condensed phase matter, like molecular liquids or molecular crystals. Whereas we have to our disposition quite reliable quantum chemical methods to describe molecular observables of the isolated species, or small aggregates, there are no standard methods for the study of the condensed phase intermolecular interactions on the electronic structure.

In the recent years a powerful technique, the periodic Hartree-Fock method has been developed for the description of three-dimensional solids in terms of Bloch-orbitals. It offers a unique tool to study especially covalent and ionic crystals. Nevertheless it has its practical limitations for systems with large unit cells (more than about 25 atoms). This limit is quite easily attained in the case of molecular crystals.

An alternative approach is offered by model Hamiltonian methods, where the building blocks of the crystal are described by taking into account the effective potential of the surroundings. The self-consistent Madelung potential (SCMP) approach corresponds to a specific form of this effective potential, involving the electrostatic interactions, which dominate in strongly polar systems. Although the principles of the SCMP method are quite easy to understand heuristically, there are a number of approximations, which could be justified only by experience.

In the present contribution we undertake a comparison of exact periodic Hartree-Fock and approximate SCMP results on the charge distribution in polar molecular crystals. In particular the charge distribution of the water molecules has been followed in various conditions, like different ice modifications and crystal hydrates.

After a brief review of the main principles of the periodic Hartree-Fock and the SCMP approaches, the most important features of the present implementations are described. The cohesion energies, and the atomic net charges are discussed and compared for proton ordered hexagonal and cubic ice modifications, as well as for $\text{LiOH}\cdot\text{H}_2\text{O}$, lithium-hydroxide monohydrate.

2. Theory

2.1. Periodic Hartree-Fock method. The Hartree-Fock-Roothan equations for periodic systems are well known [1-2] and are recalled here only for a better understanding. The crystalline orbitals $\psi_i(\mathbf{k};\mathbf{r})$,

are represented as linear combinations of Bloch functions $\varphi_{\omega}(\mathbf{k}; \mathbf{r})$ which are expressed on the basis of atomic orbitals of each unit cell:

$$\psi_i(\mathbf{k}; \mathbf{r}) = \sum_{\omega} c_{\omega i}(\mathbf{k}) \varphi_{\omega}(\mathbf{k}; \mathbf{r}) \quad (1)$$

$$\varphi_{\omega}(\mathbf{k}; \mathbf{r}) = N^{-1/2} \sum_{\mathbf{g}} \chi_{\omega}^{\mathbf{g}}(\mathbf{r}) \exp(i\mathbf{k} \cdot \mathbf{g}) \quad (2)$$

In Eq. (2) $\chi_{\omega}^{\mathbf{g}}$ denotes the ω th atomic orbital of the unit cell characterized by the direct lattice vector \mathbf{g} . The expansion coefficients $c_{\omega i}(\mathbf{k})$ are calculated by solving for each reciprocal vector \mathbf{k} in the Brillouin zone the matrix equation:

$$F(\mathbf{k})C(\mathbf{k}) = S(\mathbf{k})C(\mathbf{k})E(\mathbf{k}) \quad (3)$$

in which $S(\mathbf{k})$ is the overlap matrix, $E(\mathbf{k})$ is the eigenvalue's diagonal matrix and $F(\mathbf{k})$ is the Fock matrix:

$$F(\mathbf{k}) = \sum_{\mathbf{g}} \exp(i\mathbf{k} \cdot \mathbf{g}) F^{\mathbf{g}} \quad (4)$$

Formally the $F^{\mathbf{g}}$ matrix elements can be written as the sum of one-electron and two-electron contributions i.e.:

$$F_{\mu\nu}^{\mathbf{g}} = \langle \chi_{\mu}^0 | \hat{F} | \chi_{\nu}^{\mathbf{g}} \rangle = H_{\mu\nu}^{\mathbf{g}} + G_{\mu\nu}^{\mathbf{g}}, \quad (5)$$

in which the dummy indices μ, ν, \dots label the atomic orbitals in the elementary cell. The one-electron contribution is the sum of the kinetic $T_{\mu\nu}^{\mathbf{g}}$ and nuclear attraction $V_{\mu\nu}^{\mathbf{g}}$ terms:

$$H_{\mu\nu}^{\mathbf{g}} = \langle \chi_{\mu}^0 | \hat{T} + \hat{V} | \chi_{\nu}^{\mathbf{g}} \rangle = T_{\mu\nu}^{\mathbf{g}} + V_{\mu\nu}^{\mathbf{g}} \quad (6)$$

The two-electron term is the sum of a Coulomb and of an exchange term:

$$G_{\mu\nu}^{\mathbf{g}} = J_{\mu\nu}^{\mathbf{g}} + K_{\mu\nu}^{\mathbf{g}} = \sum_{\lambda \sigma} P_{\lambda \sigma}^1 \sum_{\mathbf{m}} [(\chi_{\mu}^0 \chi_{\nu}^{\mathbf{g}} | \chi_{\lambda}^{\mathbf{m}} \chi_{\sigma}^{\mathbf{m}+1}) - \frac{1}{2} (\chi_{\mu}^0 \chi_{\lambda}^{\mathbf{m}} | \chi_{\nu}^{\mathbf{g}} \chi_{\sigma}^{\mathbf{m}+1})] \quad (7)$$

The \mathbf{l} and \mathbf{m} summations in eq. (7), as well as the \mathbf{g} one in eq. (4), extend in principle to the infinite set of lattice vectors, in practice the convergence in the transformation from $F^{\mathbf{g}}$ to $F(\mathbf{k})$ is very fast, and the important exchange contributions can be shown to be relatively short ranged [5]. The Coulomb series require on the contrary particular attention. the electron-electron long range terms must be combined with the corresponding electron-nuclei contributions and summed to infinity by Ewald-type techniques [6].

In an all-electron approach the electron-nuclei term takes the following form:

$$V_{\mu\nu}^{\mathbf{g}} = \int \chi_{\mu}^0(\mathbf{r}) \sum_{\mathbf{h}} \sum_A Z^A |\mathbf{r} - \mathbf{R}^A - \mathbf{h}|^{-1} \chi_{\nu}^{\mathbf{g}}(\mathbf{g}) d\mathbf{r} \quad (8)$$

where the A summation extends to the atoms of the unit cell with charge Z^A . This purely local contribution is treated by standard Ewald techniques.

The density matrix elements are obtained after integration over the volume of the Brillouin zone:

$$P_{\mu\nu}^1 = 2 \int_{BZ} d\mathbf{k} \exp(i\mathbf{k} \cdot \mathbf{l}) \sum_i c_{\mu i}(\mathbf{k}) c_{\nu i}(\mathbf{k}) \vartheta(\epsilon_F - \epsilon_i(\mathbf{k})) \quad (9)$$

In Eq. (9) the $c_{\mu i}$ are the eigenvectors elements, ϑ is the step function, ϵ_F is the Fermi level and ϵ_i is the i^{th} eigenvalue. Finally the total electronic energy per unit cell is given by:

$$\mathcal{E} = \frac{1}{2} \sum_{\mu, \nu, \mathbf{g}} P_{\mu\nu}^{\mathbf{g}} (2H_{\mu\nu}^{\mathbf{g}} + G_{\mu\nu}^{\mathbf{g}}) \quad (10)$$

The program CRYSTAL [7], developed in Torino, is an ab initio program, which allows calculations on one-, two- and three-dimensional periodic systems. In CRYSTAL standard cartesian gaussian functions are used to construct the Bloch functions defined by Eq. (2). The program requires as input the space group of the system to be calculated, the fractional coordinates of the independent atoms, the basis set and the threshold values, used for the truncation of the infinite sums. As output it provides the unit cell energy, the wavefunction and related one electron properties.

2.2. Self-consistent Madlung potential (SCMP) method [10]. In this approach one is interested in the electronic wavefunction of a molecular subunit of the crystal. Following the philosophy of the group function method of McWeeny [8-9], a set of coupled equations can be written down for the N subunits of the crystal:

$$\{H^0 + \sum_{g \neq 0} \sum_{i=1}^N \langle \mathcal{R}_i \Psi^g | \hat{V} | \mathcal{R}_i \Psi^g \rangle\} |\Psi\rangle = E |\Psi\rangle \quad (11)$$

where the summations run over the space group operations (rototranslations) \mathcal{R}_i , and over the lattice translations g . Ψ stands for the wavefunction of the motif in the $g = 0$ unit cell. H^0 is the Hamilton operator of an isolated subunit and \hat{V} is the interaction operator between different subunits. This interaction includes intersystem nuclear repulsion, electron-nuclear attraction and electron-electron Coulomb and exchange operators. These equations can be decoupled by taking into account the space group symmetry relationships between the subunit wavefunctions. By denoting $\mathcal{R}_i(g)$ the appropriate combination of the rototranslation and lattice translation, and by transforming the \hat{V} interaction operator accordingly, one can define the following effective interaction kernel function:

$$\mathcal{G} = \sum_g \sum_i \mathcal{R}_i^\dagger(g) \hat{V} \mathcal{R}_i(g) \quad (12)$$

The SCMP effective Schrödinger equation for one subsystem takes the simple form:

$$\{H^0 + \langle \Psi | \mathcal{G} | \Psi \rangle\} |\Psi\rangle = E |\Psi\rangle \quad (13)$$

This is a typical nonlinear Schrödinger equation, similar to the one used in the reaction field models of solvent effects [11].

In order to complete the SCMP theory we have to define the energy per subunit in the crystal. It is easy to see that the sum of the energies E , appearing in the effective Schrödinger equations includes the subunit-subunit interactions twice. Since the subunits are equivalent, it is straightforward to define the energy per subunit, \mathcal{E} as:

$$\mathcal{E} = \langle \Psi | \{H^0 + \frac{1}{2} \langle \mathcal{G} | \Psi \rangle\} | \Psi \rangle \quad (14)$$

As we have already shown in ref [10], the effective Schrödinger equation (13) can also be derived by applying the variational principle to the above energy functional, so by solving (13) we make stationary the approximate energy functional (14).

3. Method

3.1. Evaluation of the Coulomb series in CRYSTAL. In CRYSTAL the direct space approach is used to perform the evaluation of the necessary one and two electron integrals. The difficult problem of summing the infinite series of the electrostatic interactions appearing in Eqs. (7-8) has been solved in the following way: consider the exact charge density $\rho^e(\mathbf{r})$ of the general crystal cell labelled by

the direct lattice vector \mathbf{g} . It can be split into a model point charge distribution involving atomic multipoles up to the L^{th} order and a remainder so that the total charge density can be written as:

$$\rho(\mathbf{r}) = \sum_{\mathbf{g}} \rho^{\text{model}}(\mathbf{r}) + \rho'(\mathbf{r}) = \rho^{\text{model}}(\mathbf{r}) + \rho'(\mathbf{r}) \quad (15)$$

On the one hand the choice of $\rho^{\text{model}}(\mathbf{r})$ implies that its contribution to the potential in the reference unit cell can be efficiently computed by the Ewald technique. On the other hand the contribution of the remaining charge distribution $\rho'(\mathbf{r})$ is short ranged and therefore it is evaluated by a direct summation over the unit cells located within a sphere of finite radius the size of which can be controlled by the user.

The exchange contribution is essentially short ranged and except for metallic systems the convergence of the series is fast. A set of cutoff parameters is used to control the length of the expansion. A detailed discussion of these problems can be found in ref. [6] together with the relevant algorithms used in actual calculations.

3.2. *Ab initio* implementation of the SCMP method. For ideal periodic systems the effective interaction kernel \mathcal{G} involves a combination of infinite lattice sums of one- and two-electron integrals. Obviously the brute force calculation of such integrals would be excessively expensive. Taking into account the inherently approximate nature of the SCMP method it seemed to be desirable to introduce some conceptually simple approximations for the representation of the interaction integrals.

We have chosen the Ruedenberg approximation to decompose the multicenter integrals to at most two-center ones. As a further approximation, we developed these integrals into truncated multipole series, which leads to an atomic multipole expansion of the interaction operator. This philosophy is quite close to that of the CRYSTAL algorithm. By this procedure the Madelung problem is reduced to the analytical calculation of multipole Ewald sums involving the pairs of nuclear positions. The atomic charges in this scheme are the usual Mulliken charges.

$$Q^A = Z^A - \sum_{\mu \in A} (PS)_{\mu\mu} \quad (16)$$

while the higher atomic multipoles are defined by a generalization of the above expression. For example, the x -component of the atomic dipole moment is given by:

$$m_x^A = - \sum_{\mu \in A} (PM_x)_{\mu\mu} \quad (17)$$

In this expression the matrix M_x is defined by its elements:

$$(M_x)_{\nu\mu} = \sum_{\lambda, \sigma \in A} S_{\nu\lambda} S_{\lambda\sigma}^{-1} \{ \langle \sigma | x | \mu \rangle - S_{\sigma\mu} R_x^A \} \quad (18)$$

Actually the method is working with a multipole expansion, truncated at the dipolar level. In this dipolar approximation one obtains the following expression for the corrected SCMP Fock matrix elements, within the framework of the closed shell RHF theory:

$$F_{\mu\nu} = F_{\mu\nu}^0 - \frac{1}{2} S_{\mu\nu} (V^A + V^B) + \frac{1}{2} (\mathbf{M}_{\mu\nu} \cdot \mathbf{E}^B + \mathbf{M}_{\nu\mu} \cdot \mathbf{E}^A) \quad (19)$$

In this latter expression V^A and \mathbf{E}^A stand for the total Madelung potential and Madelung field at the atomic center A , coming from the surrounding atomic charges and dipoles. The notation $\mathbf{M}_{\mu\nu}$ stands for the three Cartesian components of the dipole moment matrices, M_x , M_y and M_z .

4. Applications

Two systems in which the water molecule can be considered as a probe have been studied. On the one hand is ice with different structures and on the other hand the hydrate of lithium hydroxide $\text{LiOH}\cdot\text{H}_2\text{O}$. The periodic Hartree-Fock calculations can be considered as a reference in order to gauge the results given by the model hamiltonian SCMP method as well as those obtained from the prototype molecule (cluster) approach.

Two structures of ice have been considered here. First, the hexagonal ice which belongs to the $P6_3/mmc$ space group as far as "half hydrogens" are disposed in position (4f) corresponding to the statistically disordered orientations of the water molecules. In actual calculation one has to consider some proton-ordered structures containing exclusively "full hydrogen" atoms. Two possible substructures of lower symmetry, hereafter referred as "ortho" and "para" hexagonal ices, were considered in this work. In the "ortho" structure the water molecule dipoles belonging to a given sheet are forming an angle of 60 degrees, while in the "para" structure the water dipoles are parallel.

	ΔE (kJ mol ⁻¹)	qO (a.u.)
"ortho" ice		
PHF (STO-3G)	-56.7	-0.464
PHF (6-31G)	-68.0	-1.109
SCMP (STO-3G)	-33.4	-0.452
SCMP (6-31G)	-115.0	-1.141
cluster (STO-3G)	-33.1	-0.419
cluster (6-31G)	-57.8	-0.895
"para" ice		
PHF (STO-3G)	-56.9	-0.464
PHF (6-31G)	-68.4	-1.109
SCMP (STO-3G)	-35.6	-0.459
SCMP (6-31G)	-119.0	-1.153
cluster (STO-3G)	-12.1	-0.456 (-0.402)
cluster (6-31G)	-23.4	-0.954 (-0.866)
"cubic" ice		
PHF (STO-3G)	-58.9	-0.467
PHF (6-31G)	-74.6	-1.114
SCMP (STO-3G)	-40.8	-0.461
SCMP (6-31G)	-127.3	-1.161
cluster (STO-3G)	-17.1	-0.458 (-0.369)
cluster (6-31G)	-17.0	-0.953 (-0.838)

Table 1. Cohesive energy ΔE and oxygen net charges qO in three model ices.

PHF: periodic Hartree-Fock; SCMP: self-consistent Madelung potential;
in parentheses: oxygen net charge of singly hydrogen bonded molecule.

The second ice modification, experimentally observed at low pressures, is known as "cubic" ice. Its structure is related to that of high-cristobalite. Half hydrogens are not necessary to preserve the full symmetry of the $P4_12_12$ space group.

The results of the calculations, performed with both STO-3G and 6-21G basis sets, are listed in Table 1. The unit cell of hexagonal ice contains twelve atoms belonging to six water molecules, which are shared between several adjacent cells. Within this arrangement it is not possible to perform a SCMP calculation in which the unit cell atoms are taken as quantum motif. In the SCMP calculations on hexagonal ice reported here, the unit cell is made of one water molecules surrounded by the atomic charges and multipoles of the remaining atoms of the cell. The two clusters simulating hexagonal ice are made of rings of six molecules extracted from the ortho and para unit cells. In the ortho cluster the six molecules are equivalent and the total dipole moment is zero. In the para cluster there are basically two kinds of molecules differing by the number of hydrogen bonds in which they are involved. The para cluster has a nonzero permanent dipole moment.

The band structures, not reported here, are very flat compared to those of covalent or metallic crystals (cf. ref. [12]). This means that the k -space integration appearing in the evaluation of the density matrix (Eq. 9) can be safely approximated by a discrete sum over the occupied states. In other words this means that a localized approach is physically consistent for such systems.

The interaction energy computed by the periodic Hartree-Fock method contains the electrostatic, induction and exchange contribution but does not account for the dispersion forces. Nevertheless, the cohesive energies calculated at this level with both basis sets have the correct order of magnitude (the experimental binding energy is about 55 kJ mol^{-1}) [13-16] and are almost structure independent though the cubic ice appears to be overstabilized. They are noticeably underestimated in cluster calculations and show large discrepancies between the three structures which can be interpreted in terms of non-equivalence of the water molecules due to the limited size of the cluster. The SCMP method does not take into account the repulsive forces and therefore the interaction energy is expected to be overestimated with respect to the periodic Hartree-Fock value. In fact, this occurs with the 6-31G basis set, while STO-3G calculations seem to underestimate the cohesion energy. Nevertheless, it is remarkable, that the order of stability of the three ice modification is the same in both the PHF and SCMP calculations.

	periodic HF	SCMP	cluster
q_{Li} a.u.	0.623	0.700	0.860
$q_O(OH)$	-0.636	-0.800	-0.693
$q_H(OH)$	0.068	0.100	-0.084
$q_O(H_2O)$	-0.504	-0.466	-0.619
$q_H(H_2O)$	0.225	0.233	0.227

Table 2. Charge distribution in of lithium hydroxide monohydrate. The charges on the LiOH moiety were kept fixed in the SCMP calculation.

The meaning of the atomic net charges in the sense of Mulliken is physically questionable and it is known that these quantities are very sensitive to the nature of the basis set. In the case of this series of calculations in which the same basis set has been used to compute the same system at different levels of representation of the surroundings it appears to be a suitable tool for the analysis of the wavefunctions. Like the cohesive energy, the oxygen net charge is essentially basis set dependent. The periodic Hartree-Fock and SCMP values are in excellent agreement while the finite size of the clusters is responsible for the underestimation of this quantity.

The structure of $\text{LiOH}\cdot\text{H}_2\text{O}$ has been determined from X-ray experiments by Hermansson and Thomas [17], the unit cell belongs to the $C2/m$ space group and contains 24 atoms. The nearest Li^+ and OH^- ions are located in different unit cells and therefore the SCMP study can not be directly performed on the unit cell. The SCMP calculations were done for the H_2O motif in the fixed charge distribution of the Li^+ and OH^- ions, represented by point charges. The rather large size of the unit cell does not make possible to carry out periodic Hartree-Fock calculations of this system with a split valence basis set and therefore only STO-3G results are reported in table 2.

The agreement between the PHF and SCMP charge distributions is less spectacular than for the ice modifications. This slight discrepancy can be attributed to the fact, that the LiOH part of the system was not relaxed. In the SCMP model the charge transfer between the H_2O and LiOH motifs is not possible, while it seems to be non-negligible according to the PHF (and cluster) calculations.

Acknowledgements. The authors wish to thank Prof. C. Pisani for his interest and fruitful discussions. The Direction Scientifique of the Chemistry Department of CNRS is gratefully acknowledged for providing free CPU on the VP200 machine of CIRCE (CNRS) on which the calculations have been carried out.

REFERENCES

- [1] G. DEL RE, J. LADIK and G. BICZÓ, 1967, *Phys. Rev.*, **155**, 997.
- [2] J.M. ANDRÉ, L. GOUVERNEUR and G. LEROY, 1967, *Int. J. Quant. Chem.*, **1**, 427.
- [3] J. DELHALLE, L. PIELA, J.L. BRÉDAS and J.M. ANDRÉ, 1980, *Phys. Rev.*, **B**, **22**, 6254.
- [4] A. BEYER and A. KARPFFEN, 1982, *Chem. Phys.*, **64**, 343.
- [5] C. PISANI, R. DOVESI and C. ROETTI, 1988, *Hartree Fock ab initio treatment of crystalline systems, Lecture notes in Chemistry, vol 48*, (Springer Verlag).
- [6] R. DOVESI, C. PISANI, C. ROETTI and V.R. SAUNDERS, 1983, *Phys. Rev.*, **B**, **51**, 195.
- [7] C. PISANI, R. DOVESI, C. ROETTI, M. CAUSÁ and V.R. SAUNDERS, 1988, CRYSTAL QCPE program 577.
- [8] R. McWEENY, 1960, *Rev. Mod. Phys.*, **32**, 355.
- [9] R. McWEENY and B. T. SUTCLIFFE, 1969, *Methods of Molecular Quantum Mechanics*, (Academic Press, London).
- [10] J. G. ÁNGYÁN and B. SILVI, 1987, *J. Chem. Phys.*, **86**, 6957.
- [11] O. TAPIA, 1982, in: H. Ratajczak and W.J. Orville-Thomas (eds) *Intermolecular interactions Vol 2.*, Chapter 2., (Wiley, Chichester).
- [12] J.G. PASTORI PARRAVICINI and L. RESCA, 1973, *Phys. Rev.*, **B7**, 3009.
- [13] D. E. EISENBERG and K. KAUZMANN, 1969, *The Structure and Properties of Water*, (Oxford University Press, Oxford).
- [14] J.D. MORSE and S.A. RICE, 1982, *J. Chem. Phys.*, **76**, 650.
- [15] E.R. DAVIDSON and K. MOROKUMA, 1984, *J. Chem. Phys.*, **81**, 3741.
- [16] D.P. SANTRY, 1972, *J. Am. Chem. Soc.*, **94**, 8311.
- [17] K. HERMANSSON and J.O. THOMAS, 1982, *Acta Cryst.*, **B38**, 2555.

THEORETICAL STUDY OF INTERMOLECULAR PROTON TRANSFER IN GLYCINE

J. Langlet, J. Caillet, E. Evleth, and E. Kassab

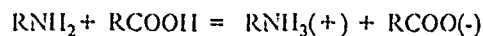
Dynamique des Interactions Moléculaires, Université Pierre et Marie Curie, Paris VI, Tour 22, 4 Place Jussieu, 75252 Paris (France)

SUMMARY

SCF 6-31G* level calculations are presented on the structures and relative energies of the neutral and zwitterionic forms of glycine and its monohydrate. Major differences are found between those structures reported here and previous published work.

INTRODUCTION

The relative energies of the neutral (GN) and zwitterionic (GZ) forms of glycine (G) and other amino acids have long fascinated experimental and theoretical chemists. The fact that the GN form is more stable than the GZ form in the gas phase can be rationalized from the relative gas phase proton affinities (PA) of alkylamines ($\text{CH}_3\text{NH}_2 = 214 \text{ kcal/mol}$) and alkylcarboxylate anions ($\text{CH}_3\text{CO}_2^- = 349 \text{ kcal/mol}$) (refs. 1,2). Without a coulombic correction term for the attraction between the charged moieties, the following gas phase reaction would be approximately endothermic by 135 kcal/mol.



However, in the case of GZ, the separation of the positive and negative charges (ca 2.4-2.7 ang) would produce a stabilizing energy on the order of between 120-135 kcal/mol. The net results is that the GZ form is predicted to be less stable than the GN form by 0-15 kcal/mol. This estimate is quantitatively unreliable but indicates that the GZ form is probably less stable than the GN form by not more than 30 kcal/mol in the gas phase. This energy difference is inverted in polar solvent solutions or in those crystal configurations which stabilize the GZ more than the GN form (ref. 3,4). In water, an estimate of the free energy difference can be made from the pK_a differences between protonated alkylamines and neutral carboxylic acids (4-5 units). This estimate indicates that at neutral pHs, the zwitterionic form (GZ) dominates in water. However, more than one species exists in appreciable concentrations than can exist in the gas phase. These species are, respectively, the carboxylate and amino protonated species: G^- , $(-)\text{OOC-CH}_2\text{-NH}_2$, and GH^+ , $\text{HOOC-CH}_2\text{-NH}_3^+$.

The problem of theoretically treating the energy difference between the GZ and GN species in the gas and solution phase has been approached by various investigators in recent years. We won't cite earlier CNDO-INDO work since these methods were intrinsically incapable of yielding reliable quantitative work. More reliable semiempirical

methods have now been developed which will certainly be used in molecular modeling schemes, e.g. AM1 (ref. 5). However, it is still not clear just how accurate these will be for modeling water-neutral or water-charged molecule interactions (ref. 6). Most of the recent studies on GN and GZ have, however, been done using *ab initio* methods (refs. 7-10) even though a recent INDO study is reported (ref. 11) on a related system. Although the expression "*ab initio*" gives the impression of quantitative reliability, computations on this size of system can not be done at the near Hartree-Fock limit level and fully corrected for correlations effects. *Ab initio* computations on polyhydrated GN and GZ species require use of modest basis sets and neglect of correlation energy corrections. Since the GZ-GN gas phase energy difference is experimentally unknown, the usual practice of increasing the basis set of the computations until one obtains what is considered a "correct" result can not be followed. Even so, it is expected that smaller basis set levels (STO-3G, DZ) would not be particularly satisfactory and one might begin to approach something reasonable at the DZP level. It is known that at the 6-31G* SCF level ion-molecule interactions of small systems accidentally mimic larger basis set, correlationally corrected calculations (ref. 12). At the 6-31G SCF level, the computation of GZ and GN with a number (1-5) of molecules of water was found technically possible in our work (ref. 13). At the 6-31G* SCF level, smaller levels of hydration can be computed.

With regard to modeling GZ and GN in the presence of a large number of water molecules, the usual (refs. 14,15) but not exclusive practice (ref. 10) has been to do Monte Carlo modeling. Monte Carlo modeling schemes, however, are based on pair-wise parameters obtained from GZ and GN single water interaction potentials. These potentials are usually not tested against direct *ab initio* calculations for small water cluster systems in which *ab initio* calculations are now possible. The work presented here and to be published elsewhere (ref. 13) will lead to such comparisons. One of the factors discussed here is whether earlier calculations, in which elaborate optimizations could not be done, obtained the correct structures and interaction energies for water-GN and GZ interactions.

METHODOLOGY

The computations reported here were carried out with the program MUNGAUSS (ref. 16) using the IBM 3090 at CIRCE, Orsay, France. All computations were carried out at the SCF level and geometries were gradient optimized under geometry C, restraint.

RESULTS AND DISCUSSION

The relative energies of the GN and GZ forms are shown in Fig. 1. In the deprotonated form, G(-), the most stable conformer is actually an unshown staggered configuration which is more stable by 13.4 kcal/mol than the eclipsed form shown in Fig. 1. This relatively large conformational energy difference probably results from large repulsions between the COO(-) moiety and the nitrogen lone pair in the eclipsed conformer in comparison with the staggered form. In this latter conformer both H-H repulsions are reduced and oxygen-NH

protons are in an attractive configuration. The energy value shown in Fig. 1 for the G(-) form (373 kcal/mol) represents its proton affinity (PA) relative to GN. The unshown staggered form of G(-) has a PA of about 360 kcal/mol which is close that above cited value for carboxylate anions of about 350 kcal/mol. However, quantitative accuracy in the estimation of PAs can not be attained at this basis set level, especially in the absence of correlation and zero point energy corrections.

With regard to various GZ conformers, the GZ shown in Fig. 1 is the most stable and displays an unusually short single hydrogen bond (1.547 ang.) between the ammonium and carboxylate moieties. This was also found in previous work (ref. 7). An unshown staggered conformer having a bifurcated $\text{-HNH}_2\cdots\text{O-C}$ double hydrogen bond is only 3 kcal/mol less stable. This small energy difference indicates that rotation about the -NH_3^+ bond might be easier than expected (see refs. 6,17), especially for a multihydrated species.

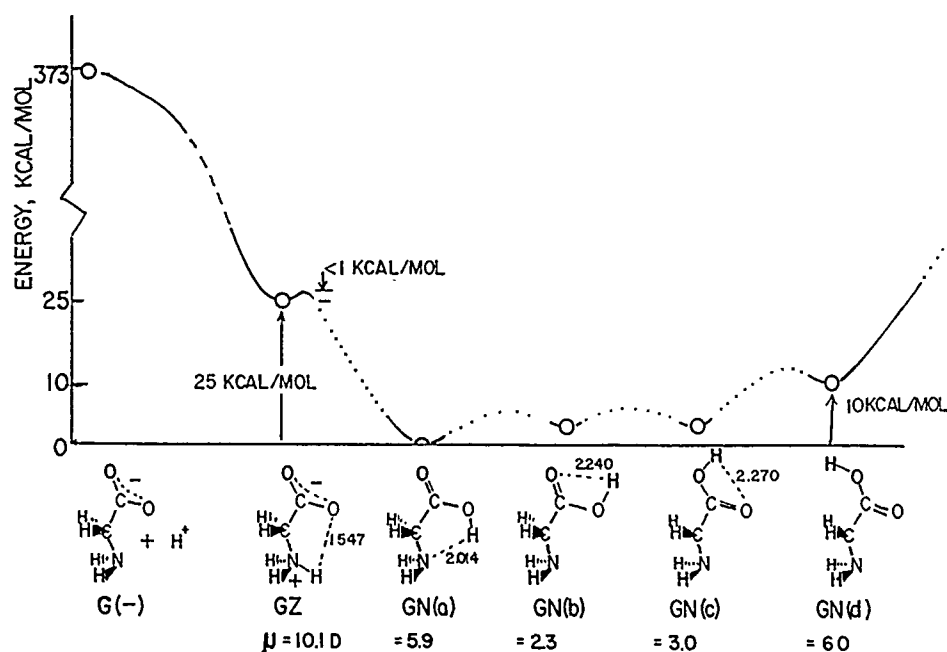


Fig. 1 Schematic representation of the proton transfer coordinate for the molecule glycine in its various forms. The energies shown were obtained at the SCF 6-31G* level, optimized in C_s symmetry. The energy of the the most stable GN (a) form is -282.82596 au. The values of the dipole moments are shown in debyes.

The GZ structure displays a minor minimum on the hypersurface at this basis set level. We found no minimum using basis sets smaller than 6-31G (STO-3G, 3-21G, 4-31G) although previous work (ref. 7) indicated one. In any case, the barrier for proton transfer is

computed to be less than 1 kcal/mol in the gas phase. Therefore, the existence of this species in the gas phase is dubious. A possible verification of the energy difference between the GZ and GN species would be to measure the gas phase activation energy for deuterium scrambling in the species $\text{ND}_2\text{-CH}_2\text{-COOH}$. However, the synthesis of this molecule in an isotopically pure form would be difficult.

The relative stabilities of the GN conformers are shown in Fig. 1. These forms are structurally related to the most stable GZ form having an intramolecular hydrogen bond. The GN(a) conformation is the most stable having a slightly shorter intramolecular H-bond, 2.014, than found in the GN(b) and GN(c) forms. These latter two structures are only 2.9 and 2.4 kcal/mol less stable than GN(a). The GN(a) form is 25 kcal/mol more stable than the GZ form. This energy difference has been estimated previously in the literature, in some cases without full geometry optimization. For instance, at the optimized 6-31G level this value is 22 kcal/mol whereas a previous non-optimized value was 43 kcal/mol (ref. 8). On the other hand, the optimized 4-31G value is 29 kcal/mol (ref. 7). These comparisons show the importance of geometry optimization in estimating the GN-GZ energy difference. The effect of correlation on this energy difference is currently under study. However, based on previous experience we would estimate that the GZ-GN energy difference will be in the 20 kcal/mol region at the best level of ab initio treatment.

We have not determined the barrier in rotation about the C-OH bond for the conversion of GN(a) to GN(b) nor the conversion between GN(b) and GN(c). The final structure, GN(d) is the least stable by 10.4 kcal/mol. This form does not have any intramolecular $\text{OH}\cdots\text{O}=\text{C}$ hydrogen bonding to stabilize its conformation.

With regard to the hydrated forms of the GN and GZ space does not permit a discussion of the structures of complexes containing up to 5 molecules of water. The two most stable structures of the monohydrate of the GZ form are CZ1 and CZ2 (Fig. 2). These structures have nearly the same energies. The first structure, CZ1, has an unusual water bridged configuration between the ammonium and carboxylate moieties. This hydrated structure still maintains the intramolecular hydrogen bond which occurs in water free GZ. This double bridged structure is unreported in previous calculations (e.g. ref. 9) in which a bifurcated complexation of water with the carboxylate function is found. However, our level of theoretical treatment finds that latter structure is only 1 kcal/mole less stable than the CZ1 structure. The second most stable structure, CZ2, has only a single bonded water. In any case, these structures are of comparable stabilities and their relative energies may be sensitive to the basis set level used beyond the DZP (e.g. 6-31 + G^* , in which diffuse functions are employed which would better describe the anionic portion of GZ).

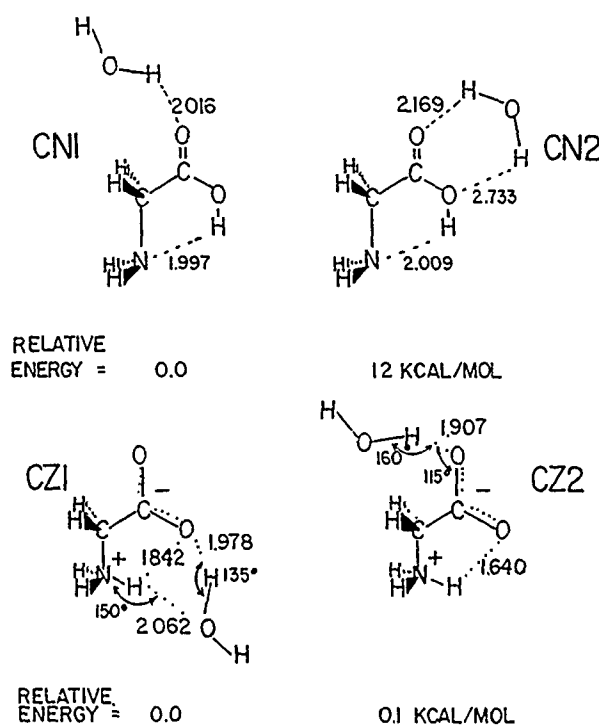


Fig. 2 The SCF level 6-31G* structures of the two most stable forms of the monohydrated GZ and GN structures. Only the hydrogen bonded structural components are shown.

With regard to the GN monohydrates, the two most stable structures were also determined (Fig. 2). In this case the most stable is singly hydrogen bonded, CN1, which is structurally related to CZ2. The CN2 configuration is bifurcated with respect to the complexation of water to the $-\text{COOH}$ moiety. Energetically, the most important feature of the calculation is that the GZ-GN energy difference decreases from 25 kcal/mol in the non-hydrated form to 21 kcal/mol in the monohydrate. Therefore, at the monohydrate level, the GN form is still more stable than the GZ. As expected, this energy difference progressively decreases with the number of water molecules. Our preliminary analysis shows that with 4-5 molecules of water the hydrated GZ and GN energies become nearly the same. (ref. 13).

This work was sponsored by a IBM-CNRS "Modélisation Moléculaire" project. The authors thank these sponsors for the work presented here.

REFERENCES

- 1 D. Aue and M.T. Bowers, Stabilities of Positive Ions from Equilibrium Gas-Phase Basicity Measurements, Gas Phase Ion Chemistry, Vol. 1 Academic Press, N.Y., 1979, pp. 2-47.
- 2 J.E. Bartmess and R.T. McIvers, Jr., Gas Phase Acidity Scale, Gas Phase Ion Chemistry, Vol. 2, Academic Press, N.Y., 1979, pp. 88-119.

- 3 R. Destro, R.E. Marsh, and R. Biacchi, A Low Temperature (23 K) Study of L-Alanine, *J. Chem. Chem.*, 92 (1988) 966-973.
- 4 R. Destro, R. Biacchi and G. Morosi, Electrostatic Properties of L-Alanine from X-Ray Diffraction at 23 K and ab Initio Calculations, *J. Phys. Chem.*, 93 (1989) 4447-4457.
- 5 M.J.S. Dewar, E.G. Zoebisch, E.F. Healy, and J.J.P. Stewart, AM1: A New General Purpose Quantum Mechanical Method, *J. Am. Chem. Soc.*, 107 (1986) 3902-3909.
- 6 E. Kassab, E. M. Evleth, and Z.D. Hamou-Tahra, Theoretical Characterization of the Rotational Motion of NH_2 in Water Clusters, *J. Am. Chem. Soc.*, in press.
- 7 Y.-C. Lee, M.D. Newton, S. Vishveshwara and J.A. Pople, Ab Initio Studies of the Relative Energies of Glycine and Its Zwitterion, *J. Am. Chem. Soc.*, 100 (1978) 4329-4331.
- 8 L.R. Wight and R.F. Borkman, Ab initio Self-Consistent Field Calculations of Some Small Amino Acids, *J. Am. Chem. Soc.*, 102 (1980) 6207-6210.
- 9 L. Carozzo, G. Corongul, C. Petrongolo, and E. Clementi, Analytical potentials from ab initio computations for the Interactions between biomolecules. IV. Water with glycine and serine zwitterions, *J. Chem. Phys.*, 68 (1978) 787-783.
- 10 R. Bonaccorsi, P. Palla, and J. Tomasi, Conformational Energy of Glycine in Aqueous Solution and Relative Stability of the Zwitterionic and Neutral Forms. An ab Initio Study. *J. Am. Chem. Soc.*, 106 (1984) 1945-1950.
- 11 G.D. Davis, III and J.W. King, Proton Transfer and Triol Formation in the Simulation of GABA Solvation, *Int. J. Quant. Chem. Quant. Chem. Symp.* 22 (1988) 593-600.
- 12 J.E. Del Bene, An Ab Initio Molecular Orbital Study of Hydrogen Bonding and Ion-Molecule Association in Model System for DNA Bases, *Int. J. Quant. Chem. Quant. Biol. Symp.* 15 (1988) 119-135.
- 13 E. Kassab, J. Langlet, J. Caillet, E. Evleth, to be published.
- 14 S. Romano and E. Clementi, Monte Carlo Simulation of Water Solvent with Biomolecules. Glycine and the Corresponding Zwitterion, *Int. J. Quant. Chem.* 15 (1978) 839-850.
- 15 G. Alagona, G. Ghio, and P. Kollman, Monte Carlo Simulations Studies of the Solvation of Ions. 2. Glycine Zwitterion, *THEOCHEM* 166 (1988) 395-392.
- 16 R.A. Poirier, Department of Chemistry, Memorial University of Newfoundland, St. Johns.
- 17 C.L. Perrin, Proton Exchange in Amides: Surprises from Simple Systems, *Accts. Chem. Res.*, 22 (1989) 268-275.

DOUBLE PROTON TRANSFER STUDIES IN CARBOXYLIC ACID DIMERS

H. CHOJNACKI and M.J. PYKA

Institute of Organic and Physical Chemistry (I-4), Wyb. Wyspiańskiego 27, 50-370 Wrocław,
Poland

SUMMARY

The potential energy curves for the double proton transfer in propionic, butyric, valeric, benzoic and o-chlorobenzoic acids have been calculated by the modified all-valence INDO approach. The calculation has been done for the isolated centrosymmetric dimers and the same systems in the crystal lattice. The role of substituents in the potential shape is discussed.

INTRODUCTION

Formic and acetic acids occur in the crystalline state as hydrogen bonded linear polymers, whereas higher of the $\text{CH}_3(\text{CH}_2)_n\text{COOH}$ homologs exist in the solid as centrosymmetric dimers (refs. 1-5). The dimeric systems have an important common feature, i.e., eight-membered ring consisting of the two $\text{O}-\text{H} \cdots \text{O}$ hydrogen bridges supplemented with the two additional carbon atoms. The systems are relatively simple and can be considered as models for the double proton transfer studies in other more structurally complicated and biologically important compounds.

It is preferentially assumed that in the gaseous state the protons can displace within the hydrogen bonds between the two symmetrical potential minima of equivalent tautomers. In a condensed medium, however, because of the interacting with the environment, the initial and final states may be trapped in some configuration. Thus, the role of the crystal field in the mechanism of this reaction may be significant but still not clearly understood. In such a situation a comparison of the potential energy curves for the proton displacement in the isolated gaseous dimers and those in the crystal might be interesting.

THEORETICAL OUTLINE

The charge distribution over the molecule under consideration in the crystal lattice has been evaluated within the iterative procedure. First, we calculate atomic charges for the free molecule and then take into account potential created by the crystal lattice. In the next step the new charges considering previously calculated potentials are found. This procedure is being repeated up to obtain stable charges with a given accuracy.

The atomic net charges have been evaluated by the quantum mechanical modified all-valence INDO method (ref. 6). However, the calculation of electrostatic potentials is more complicated because of the summation over the whole crystal lattice. This is infinite and slowly convergent series

$$V_{iM} = \sum_{j \notin M}^{\infty} \frac{q_j}{r_{ij}} \quad (1)$$

where

V_{iM} is the electrostatic potential on i -th atom of the M -th molecule

q_j denotes the j -th atom charge

r_{ij} is the respective interatomic distance.

The summation in Eq. (1) may be partitioned into the following three steps

$$V_{iM} = \sum_{k=1}^{\text{lattice}} \sum_{m \in k}^{\text{cell}} \sum_{\substack{j \in m \\ m \neq M}}^{\text{molecule}} \frac{q_j}{r_{ijmk}} \quad (2a)$$

where

index j denotes the summation over one molecule

index m denotes the summation over all molecules in unit cell

index k denotes the summation over all cells in lattice.

Changing the summation order in Eq. (2a), we obtain

$$V_{iM} = \sum_{\substack{j \in m \\ m \neq M}}^{\text{molecule}} q_j \underbrace{\left(\sum_{k=1}^{\text{lattice}} \sum_{m \in k}^{\text{cell}} \frac{1}{r_{ijmk}} \right)}_{M_{ij}} \quad (2b)$$

The inconvenient infinite sum is independent of atomic charges and it can be calculated by using the Ewald method (refs. 7-8). In order to evaluate the electrostatic potentials in the above described iterative procedure, we have to do only the following summation

$$V_{iM} = \sum_{\substack{j \in m \\ m \neq M}}^{\text{molecule}} q_j M_{ij} \quad (2c)$$

RESULTS AND DISCUSSION

The results of our calculation (Table 1) show that the barrier height is strongly dependent on the substituents. Mainly, benzene and *o*-chlorobenzene rings decrease the barrier leading to the symmetrical potential in the former case. No symmetrical potential for the double proton displacement was found for other systems under consideration. On the other hand, the aliphatic chain length seems do not influence the barrier. These conclusions are based

on the assumption that the hydrogen bond length is almost the same in all systems under consideration.

Table 1: Molecular parameters and barrier height values [eV] for carboxylic acid dimers.

ACID DIMER	PROPIONIC Ref. 1	BUTYRIC Ref. 2	VALERIC Ref. 3	BENZOIC Ref. 4	O-CHLORO -BENZOIC Ref. 5
H-B LENGTH	2.64	2.62	2.64	2.64	2.63
DIFFERENCE*	0.09	0.12	0.09	0.01	0.09
GASEOUS	1.42	1.10	1.30	0.28	0.83
CRYSTAL	1.42	0.95	1.35	0.32	0.79

* Difference [\AA] between C - O, and C = O bond lengths

The potential energy curve for the benzoic acid dimer seems to be nearly symmetric and the lowest of the all calculated systems. This result is in agreement with the latest Hochstrasser and Trommsdorff experimental results (ref. 9) showing that the acid protons are delocalized within the approximately symmetric double-well potential (Fig. 1).

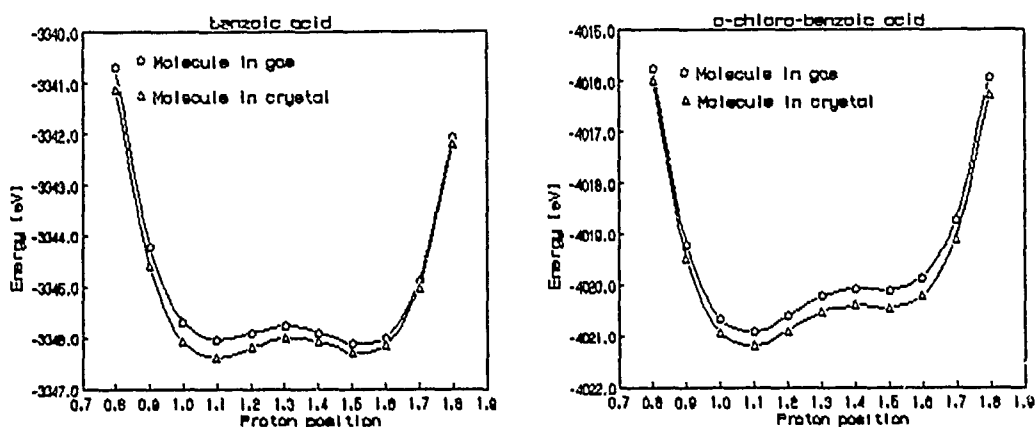


Fig. 1: Potential energy curves for the double proton transfer of the benzoic acid and o-chloro-benzoic acid dimer.

In fact, it was the first observation of the protons' delocalization for carboxylic acid dimers in the condensed phase which seems to be in a good correlation with the difference in the C-O - and C=O bond lengths (Table 1). The potential energy maps for the double proton transfer

in benzoic gaseous and crystal Fig. 2 acid dimer shows that the two hydrogen bond protons do not move independently. The correlated double proton transfer has also been confirmed by calculational results in our previous studies (ref. 10).

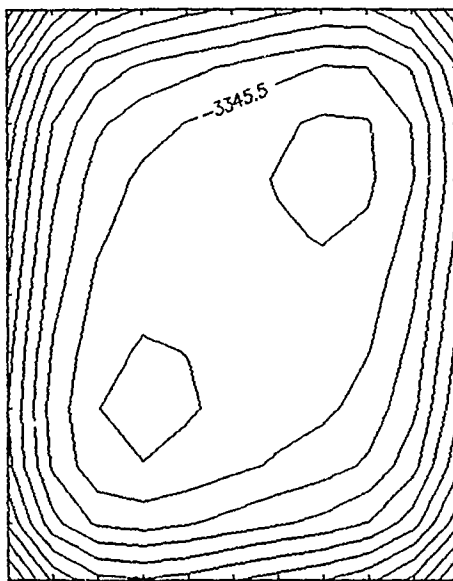


Fig. 2: Potential energy [eV] map for the double proton transfer in the crystalline benzoic acid.

The potential energy curve for the torsional rotation of the eight-membered ring suggested by Furić (ref. 11) has been calculated for benzoic and o-chloro-benzoic acids (Fig. 3). According to the Grabowski and Krygowski (ref. 12) calculations, this kind of tautomeric transformation in the crystal lattice seems to be possible. In our case, however, as the potential is positive for some rotation angles, the Furić mechanism of tautomerization should be excluded.

It should be noted that the method of our calculations may be much too simplified for quantitative considerations, however, it seems to be conclusive for comparative discussion. Furthermore, the molecular geometry optimization may influence the barrier height to a great extent. Anyway, our experience (ref. 13) says that the results of our calculations may be useful for qualitative considerations including experimental investigations of these important tautomerization reactions.

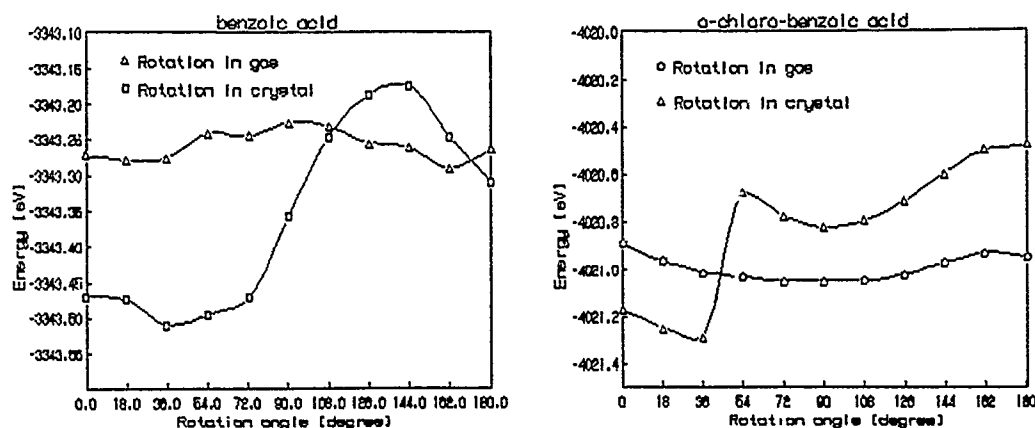


Fig. 3: Potential energy curves for hindered rotation of the eight-membered ring in the gaseous and crystalline benzoic acid and o-chloro-benzoic acid dimer.

In general, our calculations strictly point out that the crystal lattice effects seem to lower the potential barrier for linear hydrogen bonded systems (ref. 14), whereas no essential changes were found for carboxylic acid dimers. Then, in agreement with Nakamura and Hayashi (ref. 15) these effects reduce the barrier height by ca. 15% and seem to be unable to explain experimental results of the low potential energy barrier of the 1 kcal/mol for the crystalline benzoic dimer (ref. 16) whereas the calculated activation energy was estimated to be 5 kcal/mol (ref. 17) or more.

REFERENCES

- 1 F.J. Strieter, D.H. Templeton, *Acta Crystallog.*, 15(12) (1962) 1233-1239.
- 2 R.F. Scheuerman and R.L. Sass, *Acta Crystallog.*, 15(12) (1962) 1244-1247.
- 3 F.J. Strieter and D.H. Templeton, *Acta Crystallog.*, 15(12) (1962) 1240-1244.
- 4 R. Feld and M.S. Lehmann, *Z. Krist.*, 157 (1981) 215-231;
G.A. Sim, J.M. Robertson and T.H. Goodwin, *Acta Crystallog.*, 8 (1955) 157-164.
- 5 G. Ferguson and G.A. Sim, *Acta Crystallog.*, 14(12) (1961) 1262-1268.
- 6 J. Lipiński, *Int. J. Quant. Chem.*, 34(5) (1988) 423-435.
- 7 M. Catti, *Acta Crystallog.*, A34(6) (1978) 974-979.
- 8 L. Komorowski, J. Lipiński, P. Misiak, M.J. Pyka, *Sci. Pap. Inst. Inorg. Chem. Met. Rare Elements*, No 57, Ser. Conferences No 12, Wrocław 1988, 54-64.
- 9 R.M. Hochstrasser and H.P. Trommsdorff, *Chem. Phys.*, 115(1) (1987) 1-6.
- 10 H. Chojnacki and Z. Laskowski, *J. Mol. Struct.*, 123(1-2) (1985) 121-127.

- 11 K. Furić, Chem. Phys. Lett., 108(5)(1984)518-521.
- 12 S.J. Grabowski and T.M. Krygowski, Chem. Phys. Lett., 151(4-5) (1988) 425-427.
- 13 M.J. Pyka and H. Chojnacki, in preparation.
- 14 H. Chojnacki and Z. Laskowski, unpublished results.
- 15 R. Nakamura and S. Hayashi, J. Mol. Struct., 145(3-4) (1986) 331-339.
- 16 S. Nagaoka, T. Terao, F. Imashiro, A. Saika and N. Hirota J. Chem. Phys., 79(10) (1983) 4694-4702.
- 17 S. Hayashi, J. Umemura, S. Kato and K. Morokuma, J. Phys. Chem., 88(7) (1984) 1330-1334.

DISCUSSION

PERAHIA - Do you think the role of conformational relaxation may be important to lower down the barrier of double proton transfer associated with the 8 membered ring rotation ?

CHOJNACKI - I am convinced that the role of relaxation is important to change the potential energy curve for the proton transfer reactions. However, in our comparative studies, we have performed the scaling for the total energy, and the results should be still conclusive.

ANGYAN - 1. The self-consistent Madelung Potential approach you use should be adapted very carefully with respect to the energy expression. In effect, the perturbation in the Hamiltonian due to the crystal field is :

$$\hat{V}(r) = \int \hat{\rho}(r') G(r,r') < \psi | \hat{\rho}(r') | \psi > dr'$$

is of non-linear nature, due to the appearance of the expectation value. This non-linear character is reflected by the fact that the energy expression (energy per molecule in the crystal) is :

$$E = < \psi | H^0 + 1/2 \int dr V(r) \rho(r) | \psi >$$

which is different from your energy expression by a factor of 1/2 before the interaction term (see J.G. Angyan & B. Silvi, J. Chem. Phys. 86 (1987) 6957).

Although your wavefunction should be correct the energies should be corrected by the above-mentioned factor. This may affect the results, and the conclusions concerning the negligible effect of the crystal field may be revised.

2. Your model corresponds to a simultaneous proton transfer in the crystal. Physically it is perhaps more reasonable to consider such a process as a "defect" formation, where only one pair of H-bonds is reorganizing and creates an ionic defect. A similar model was treated by us recently on the proton conduction in oxonium perchlorate. Can you comment on the difference between the "simultaneous" and "point-defect" proton transfer models.

CHOJNACKI - 1. It may be not clearly seen from the text, but in fact we have taken into account the non-linear factor of $1/2$ in our computer program for the energy evaluation.

2. Our calculations on the defect mechanism are in progress. However, it seems now that the proton transfer reaction is the correlated process.

MODELLING OF THE DISULFIDE BRIDGE IN PROTEINS : *AB INITIO*-CI STUDIES OF S_2H_2 AND $S_2(CH_3)_2$

Michel LOOS

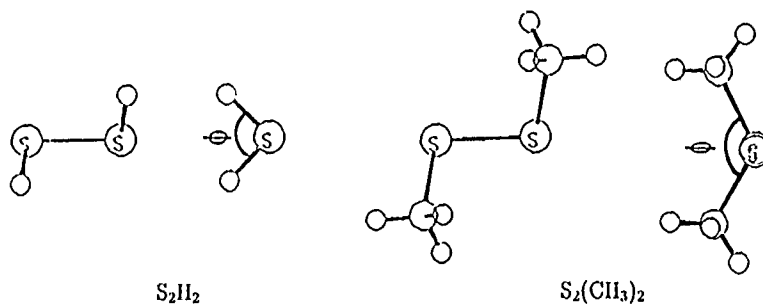
Laboratoire de Chimie Théorique, U.A. 510 CNRS, Université de Nancy-I, BP 239, F-54506
Vandœuvre-les-Nancy, Cedex, France.

SUMMARY

We report here the potential surfaces around the S—S bond for S_2H_2 and for $S_2(CH_3)_2$, using *Ab initio*-CI calculations with a 4-31G basis set.

INTRODUCTION

The conformation of the disulfide bond in proteins plays a key role in their biological properties[1]. Much theoretical work has been done concerning the rotational barriers about the S-S bond[2,3]. We calculated the potential surface in the fundamental state and in the lowest excited state in order to evaluate the possibility of radiation damage to the proteins.



We used two models for the disulfide bond : the S_2H_2 (Cf. Fig 1a) and $S_2(CH_3)_2$ (Cf Fig 1b) systems and we report here the torsional potential surfaces around the S-S bond for both of them in the fundamental and lowest excited state.

METHOD

Ab initio SCF calculations have been carried out using the HONDO program [4] at the 4-31G*

basis set level. The coefficients of the d-type orbitals which were optimized on H_2S are given in table I.

	Expt.	Coef.
d functions	0.818916	0.357851
	0.262078	0.759561

Table I

All geometries were gradient optimized at the SCF level with one frozen parameter (θ) and the optimization was stopped when the gradients in internal coordinates were less than 5.10^{-4} .

CI calculations including up to the third level of excitation were then carried out using the same program both for the fundamental state and the lowest excited state.

Every potential curve was fitted with a function of the type :

$$A + B * \cos(\theta) + C * \cos(2 * \theta) + D * \cos(3 * \theta) \quad (1)$$

used in most molecular mechanics programs.

RESULTS

These calculations give us the rotational barrier of both compounds : respectively $\Delta E_1 = 11.5$ kcal/mol and $\Delta E_2 = 5.7$ kcal/mol for the eclipsed and planar forms of $\text{S}_2(\text{CH}_3)_2$ and $\Delta E_1 = 9.4$ kcal/mol and $\Delta E_2 = 7.6$ kcal/mol for S_2H_2 . These values are the same with or without CI for S_2H_2 where the correlation is very weak but for $\text{S}_2(\text{CH}_3)_2$ the differences are important. Considering the CI we get $\Delta E_1 = 9.4$ kcal/mol and $\Delta E_2 = 6.0$ kcal/mol. The reduction of the rotational barrier of the eclipsed form is due to the stabilization of some hyperconjugated forms in that conformation. The differences between our two models are now rather small.

	$\bar{E}(\text{HF})$	$\bar{E}(\text{CI})$	$\bar{E}(\text{CI})$ excited
S_2H_2 (cis)	-795.3661817	-795.6327498	-795.2627430
S_2H_2 (90°)	-795.3811301	-795.6473034	-795.2762887
S_2H_2 (trans)	-795.3689627	-795.6354249	-795.3147800
$\text{S}_2(\text{CH}_3)_2$ (cis)	-873.3216696	-873.6415101	-873.2168834
$\text{S}_2(\text{CH}_3)_2$ (90°)	-873.3399725	-873.6565262	-873.2274197
$\text{S}_2(\text{CH}_3)_2$ (trans)	-873.3308588	-873.6470410	-873.2804258

Table II : Absolute energies (in Hartrees) of the critical points.

The lowest excited state corresponds to a transition of the type $\sigma_{S-S} \rightarrow \pi_{S-S}$.

We report in Fig. 2 the potential curve for the compound S_2H_2 both for the fundamental state and the lowest excited state. One can notice the modification of the minima and maxima positions between the fundamental state and the excited state. Furthermore the minimum of the excited state potential corresponds to a maximum of the fundamental state potential.

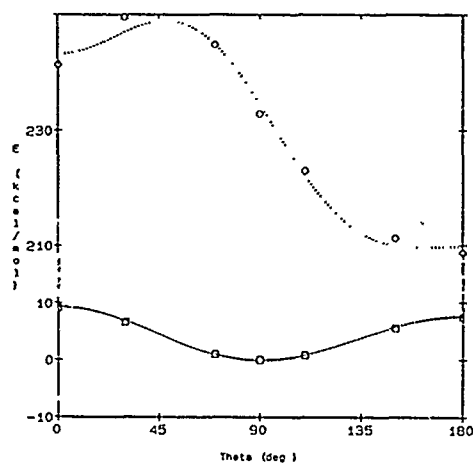


Figure 2 : Potential surface of S_2H_2

—□— : fundamental state
 ...○... : excited state

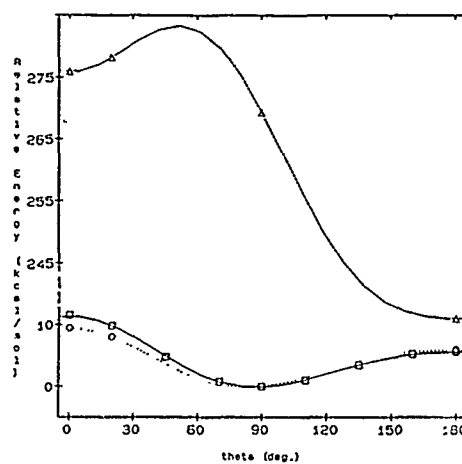


Figure 3 : Potential surface of $S_2(CH_3)_2$

—□— : fundamental state HF
 ...○... : fundamental state Cl
 ---△--- : excited state

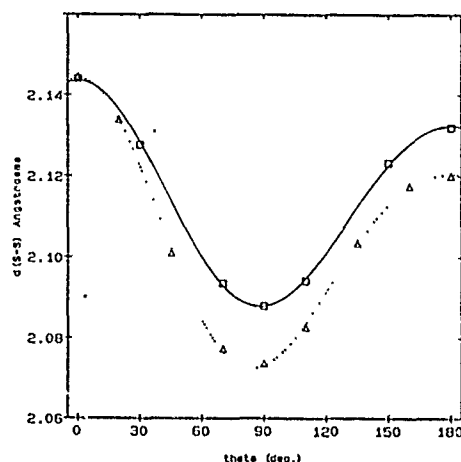


Figure 4 : Variation of the S-S distance with the torsional angle θ

—□— : S_2H_2
 ...△... : $S_2(CH_3)_2$

The same observations can be done in figure 4 concerning $S_2(CH_3)_2$.

The coefficients of equation (1) are given in table III.

	A	B	C	D
$S_2H_2 - HF$	4.19	0.65	4.25	0.21
$S_2H_2 - Cl$	4.19	0.65	4.25	0.21
$S_2(CH_3)_2 - HF$	4.22	1.92	4.29	0.97
$S_2(CH_3)_2 - Cl$	3.85	0.70	3.85	1.02

Table III : Coefficients of the fitted functions.

We also report in Fig. 3 the variation of the distance R_{S-S} with the torsional angle θ in S_2H_2 and $S_2(CH_3)_2$. This distance is strongly dependent on the torsional angle θ . Spectroscopic techniques like EXAFS which are useful to determine precisely specific distances in proteins [5] can then give us informations about that angle.

The change in the topology of the potential curves of the disulfide bonds between the fundamental and first excited state can induce a change in the absolute conformation around these bonds under radiation. Furthermore, knowing that these compounds exhibit a strong natural circular dichroism [6], one can hope to increase the relative concentration of one of the enantiomers under irradiation with circularly polarized light.

CONCLUSION

The topological change in the torsional potential between the fundamental state and the lowest excited state in disulfides can lead to an inversion of the absolute configuration of these disulfides under radiation. This inversion can be used in order to modify the relative concentration of the two enantiomers.

Acknowledgements

The author wishes to thank IBM (France) for their sponsorship.

References

- [1] M. Airla, C. Nagata
Theor. Chem. Acta 70, pp. 73-80, (1986) and *ref. therein*
- [2] W. Zhao, J. Bandekar, S. Krimm,
J. Am. Chem. Soc. 110, pp. 6891-6892, (1988)
- [3] H.E. Van Wart, L.L. Shipman, H.A. Scheraga,
J. Phys. Chem., 78, 18, pp. 1848-1853, (1974)
- [4] M. Dupuis, J.D. Watts, H.O. Villar, G.J.B. Hurst,
QCPE Bull., 8(2), pp. 79-82 (1988)
- [5] J. Goulon, M. Loos, P. Friant, M. Ruiz-Lopez, in
"Chemical Cristallography with pulsed sources of Neutrons and Synchrotron Radiation"
ed. by M.A. CARRONDO and G.A. JEFFFREY, D. REIDEL Publish. Comp.
NATO ASI Series, ISBN 90-227-2631-0
Series C : Mathematical and Physical Sciences, 221, 247-293, (1988)
- [6] M. Loos, C. Gauthier, J. Goulon to be published.

INVESTIGATION OF IONIC SOLVATION DYNAMICS BY FAR INFRARED SPECTROSCOPY

B. GUILLOT¹, Ph. MARTEAU² and J. OBRIOT²

¹Laboratoire de Physique Théorique des Liquides, Université Pierre et Marie Curie, 4, Place Jussieu, 75252 Paris Cedex 05 (France)

²Laboratoire d'Ingénierie des Matériaux et des Hautes Pressions, CNRS Université Paris-Nord, Avenue J.B. Clément, Villetaneuse 93430 (France)

SUMMARY

The far infrared spectra of alkali halide salts in solvents of high dielectric constant (CH_3OH , CH_3CN) are investigated in the frequency range 20-600 cm^{-1} . These spectra are characterized by a complex band shape exhibiting several absorption peaks. By using a chemical model of the electrolytic solution and with the help of the Mori theory, it is possible to reproduce the band shape and to identify the molecular motions at the origin of the absorption peaks.

INTRODUCTION

After one century of investigations on electrolytic solutions, the quantitative description of all the microscopic processes involved in the ionic solvation is not yet clearly established. The reason for that mainly lies in the large variety of molecular entities that constitute the solution (free ions, ion pairs of different structure, higher order ionic aggregates), each being characterized by its own state of solvation. In the last two decades many informations concerning the static and dynamic aspects of ionic solvation have been obtained with the help of conductivity and permittivity measurements, optic and magnetic spectroscopies and, more recently, by using X-ray and neutron scattering : for reviews see (refs. 1-3). From a theoretical standpoint two major advances have greatly enlarged our understanding of ionic solutions : (i) the emergence of a statistical theory of ionic solutions at equilibrium based on the correlation function formalism, (ii) the extensive use of computer simulations (MC and MD). For a report on the state of the art see ref. 4.

The present paper is concerned with the dynamics of the ionic solvation as revealed by far infrared (FIR) spectroscopy. The microscopic processes which take place in an electrolytic solution, cover a large domain of time, from 10^{-2} s to 10^{-14} s. Dielectric relaxation and conductivity measurements probe the time scale up to 10^{-10} s and give informations on the collective reorientation of ion pairs and ionic aggregates in the solution as well as on the translational diffusion of the ionic species (ref. 5). As far as solutions of low conductivity are concerned, the permittivity measurements lead to a direct access on the structure and long time dynamics of ion pairs and aggregates. On the contrary, when

solutions of high conductivity are investigated, the interpretation of the measurements is greatly complicated by the presence of a large number of free ions in the solution. However, at much higher frequencies, i.e. in the far infrared range (10^{-12} s - 10^{-14} s), the conductivity of the solution does not affect the absorption coefficient and the infrared band is then a very sensitive probe of the molecular dynamics in this time scale. This fact was recognized in the late sixties and seventies when several authors (refs. 6-13) discovered that the far infrared band of alkali metal salts in various solvents (DMSO, THF, 2-pyrrolidone, sulfolane, acetone, etc..) could be attributed to the vibration of the alkali ion in its first solvation shell. Nevertheless, the assignment of a FIR band is not obvious and strongly depends on the solvent. For example, Evans and Lo (ref. 14) found that, for tetraalkylammonium salts in benzene, the FIR band could be attributed to a cation-anion vibration. On the other hand, the FIR spectra of the alkali metal salts in solvents of higher dielectric constant were interpreted as due to the vibration of the cation in a cage formed wholly or partly, by the solvent molecules; in the latter case the anion participates to the rattling motion. However, these early studies suffer of two drawbacks. First, the assignment of the FIR band is only very qualitative since the interpretation is not based on a rigorous theoretical treatment of the band shape as it exists for Raman and infrared bands in neat liquids. Next, due to the intrinsic difficulties to record such spectra, the band shapes are poorly resolved (if published) and many spectral features could be left out.

Our purpose is to reexamine the FIR spectra of alkali halide salts in different solvents of high dielectric constant, namely, a protic solvent, methanol, and a dipolar aprotic solvent, acetonitrile. The experimental band shapes are interpreted in the framework of a time dependent correlation function formalism. The correlation functions associated with the electrolytic solution are described by a generalized Langevin theory and the parameters of the theory are estimated with the help of recent computer simulation data on ionic solutions. A quantitative description of ionic motions taking place into the solution is given and the ionic entities giving rise to specific frequency modes are identified.

EXPERIMENTAL METHOD

Absorption spectra were recorded with a Michelson interferometer, CODERG FS 2000, between 20 and 600 cm^{-1} . Since, in this frequency range, dipolar solvents such as methanol and acetonitrile exhibit rather strong absorption bands, the absorption pathlength through the sample was limited to about 50μ . The optical cell was thus made of two polyethylene windows and of a flat ring of mylar pressed between them. This cell was easily filled without bubbles by running the

liquid through two capillary plastic tubes, inserted in one of the windows. The exact sample thickness was determined, a posteriori, thanks to the interference fringes appearing in the empty cell.

All products, salts (LiI, RbI) as well as solvents (CH_3OH , CH_3CN) were used as delivered by Aldrich chemie, without further purification. In order to extract, from the total absorption spectrum of the solution, that part due to the salt-solvent interaction the spectrum of each solution was referred to that of the pure solvent, i.e. plotting the quantity $\ln I_0/I$, where I_0 and I are the transmitted intensities through the pure solvent and through the solution, respectively. However, at the relatively high concentrations of salt considered here (0.5 - 1.5 mol/kg) the number density of solvent molecules in the solution is sensibly lower than its value in the pure solvent. This fact is clearly illustrated on our absorption spectra by the existence of some small negative regions (see Fig.1). A correction of these spectra is now in progress and will appear in a further and more extended publication. Nevertheless it must be emphasized that the general spectral features examined here will remain practically unchanged.

THEORY OF THE FIR SPECTRUM OF ELECTROLYTIC SOLUTIONS

Until very recently the relationship between the electric permittivity of a polar liquid and the molecular level was the subject of a continuous debate. As a matter of fact, the frequency dependent dielectric constant is related to the correlation function of the microscopic dipole moments of the sample through an equation which depends on the geometry of the sample. This is a well known effect of the long range dipole-dipole interactions. This point has an important consequence when evaluating by computer simulation the dielectric constant of a polar medium. The shape of the basic cell and the periodic boundary conditions require to implement the relevant expression for the dielectric constant : if not, spurious results are obtained. In the case of an electrolytic solution, the problem is complicated by the electric conductivity of the medium. However, by using the linear response theory, it is possible to derive a simple expression for the permittivity, $\epsilon(\omega)$, associated with a spherical liquid sample of volume V (this system is chosen by convenience) embedded in a perfectly conducting medium (see refs. 5,15),

$$\epsilon(\omega) - \epsilon_\infty = \frac{4\pi}{3k_BTV} \left\{ \langle \mathbf{M}^2 \rangle + i\omega \langle \mathbf{M} \cdot \mathbf{J} \rangle_\omega + \langle \mathbf{M} \cdot \mathbf{J} \rangle_\omega \right\} \quad (1)$$

where \mathbf{M} takes into account all microscopic dipole moments (permanent and induced) existing in the solution, \mathbf{J} is the ionic current generated by the fraction of dissociated salt, $\langle \rangle_\omega$ is the Laplace transform of the corresponding

correlation function and ϵ_∞ is the value of the dielectric constant in the high frequency limit*. In a far infrared experiment one measures the absorption coefficient, $\alpha(\omega) = \omega \text{Im} \epsilon(\omega) / nc$, where n is the refractive index. In defining $M_{\text{ion}} = \sum q_i R_i$ (with $J = M_{\text{ion}}$), and after some algebra, the absorption coefficient of an electrolytic solution is given by,**

$$\alpha(\omega) = \frac{2\pi}{3nckTV} \omega^2 \int_{-\infty}^{+\infty} dt e^{-i\omega t} \langle (M(t) + M_{\text{ion}}(t)) \cdot M(0) \rangle \quad (2)$$

It is worthwhile to notice that in the case of a non dissociating solvent (e.g. benzene), $M_{\text{ion}} = 0$, the expression (2) becomes identical to that of a neat liquid. Nevertheless to go further in a formal analysis, it is necessary to define what we call a *chemical* model of the electrolytic solution. It is well established (ref.16) that in an electrolytic solution, free ions, a variety of ion pairs (at contact or solvent separated), triple ions and higher ionic aggregates coexist as distinct species, each endowed with its own physical properties. In highly dissociating solvents as methanol and acetonitrile, an appreciable amount of salt is dissociated. Then we postulate that the ions are either free ions, or associated in pairs; the concentration of each species being governed by the dissociation constant. For ion pairs we will assume that they can exist either as contact ion pairs (CIP), tightly bound, or as solvent separated ion pairs (SSIP), loosely bound. Moreover, we will neglect higher ionic aggregates. Thus in the framework of this model it is now possible to specify more precisely all molecular mechanisms which contribute to the absorption coefficient.

In fact, even with the small number of well defined species (free ions, ion pairs, solvent molecules), the evaluation of eqn.(2) is hardly tractable in its full complexity. Hence we choose to make the following approximations;

- a) the solvent molecules are polarized only by the coulombic charges of the ions,
- b) the polarizability α_s of the solvent molecules is assumed isotropic,
- c) the ions are not polarized by the solvent molecules.

Then, the total dipole moment due to all neutral species is given by,

$$M = M_s + M_{\text{agg}} + M_{\text{pol}} \quad ; \quad \text{with} \quad M_s = \sum_{i=1}^{N_s} \mu_i \quad (3a,b)$$

* In practice, this value is taken in the IR frequency range corresponding to the absorption of light by the intramolecular vibration modes of the molecules.

** A rigorous expression taking into account the detailed balance principle will make appear in eqn.(2), $\omega \tanh(\beta \hbar \omega / 2)$ instead of ω , and a symmetrized CF.

$$M_{agg} = \sum_{i=1}^{N_{cip}} \mu_i^{cip} + \sum_{i=1}^{N_{SSIP}} \mu_i^{SSIP} ; \quad M_{pol} = \alpha_s \sum_{\substack{\gamma=+,- \\ i\gamma,j}}^{N\gamma, Ns} E_{i\gamma j} \quad (3c,d)$$

where μ_i is the permanent dipole of the solvent molecule i , μ_i^{cip} and μ_i^{SSIP} are the dipole moment of the contact ion pair i and the solvent separated ion pair j , respectively, and $\alpha_s E_{i\gamma j}$ is the dipole moment induced on the solvent molecule j by the electric field emanating from the ion $i\gamma$. With the definitions (3a-d) and that of M_{ion} given previously, it is easy to write formally the microscopic expression of $\alpha(\omega)$; but to be useful some additional simplifications are necessary. It is impossible in this brief report to make a detailed analysis of the theory beyond this step; that will be the purpose of a next extended paper (ref.17). Nevertheless, the following comments merit attention. In highly dissociating solvents and at low concentration of salt, the free ions and the ion pairs have only solvent molecules in their first solvation shell. Since the electrostatic interactions are strongly screened by the polar solvent beyond the first solvation shell, one can reasonably assume that the molecular motions of the ions are decoupled from each other. Moreover, the ions being strongly solvated, their motions are characterized by an oscillation in the cage of the first neighbours, at short times ($\sim 10^{-14}$ - 10^{-13} s), and by a slow drift at longer time ($\sim 10^{-12}$ s). All these features can be incorporated into the theory. This leads to the following approximate formula for $G(t)$, the time CF figuring in eqn.(2),

$$G(t) = \langle M_{pol}(t) \cdot M_{pol}(0) \rangle + \langle M_{agg}(t) \cdot M_{agg}(0) \rangle + \langle M_s(t) \cdot M_s(0) \rangle \quad (4)$$

$$\begin{aligned} \langle M_{pol}(t) \cdot M_{pol}(0) \rangle = & (\alpha_s e)^2 \sum_{\substack{\gamma=+,- \\ i\gamma,j}}^{N\gamma, Ns} \left\{ \frac{u_{i\gamma j}(t) \cdot u_{i\gamma k}(0)}{r_{i\gamma j}^2(t) r_{i\gamma k}^2(0)} \right. \\ & \left. + \langle Q\gamma(t) \cdot Q\gamma(0) \rangle \frac{3(u_{i\gamma j}(t) \cdot u_{i\gamma j}(0))^2 - 1}{r_{i\gamma j}^3(t) r_{i\gamma j}^3(0)} \right\} \end{aligned} \quad (5)$$

$$\langle M_{agg}(t) \cdot M_{agg}(0) \rangle = e^2 \sum_{i=1}^{N_{cip}} \langle Q_i^{cip}(t) \cdot Q_i^{cip}(0) \rangle + e^2 \sum_{i=1}^{N_{SSIP}} \langle Q_i^{SSIP}(t) \cdot Q_i^{SSIP}(0) \rangle \quad (6)$$

$$\langle M_s(t) \cdot M_s(0) \rangle = \sum_{i,j}^{Ns} \langle \mu_i(t) \cdot \mu_j(0) \rangle_{\text{solution}} \quad (7)$$

where $r_{i\gamma j}$ is the distance between the ion $i\gamma$ and the solvent molecule j , $Q\gamma$ is the caging vibration mode of the ion $i\gamma$, Q_i^{cip} (or Q_i^{SSIP}) is the distance separating the centers of charge of the contact ion pair i (or SSIP), and μ_i is the permanent dipole moment of the solvent molecule i . In practice, what is

measured is the absorption difference between the solution and an equivalent amount of pure solvent. Thus we must subtract from $G(t)$ the CF $\langle M_s(t) \cdot M_s(0) \rangle_0$ for the neat solvent. This makes to appear a time correlation function of structural relaxation, $G_{\text{relax}}(t)$, which expresses the reorganization of the solvent molecules located just around the ionic species. As a matter of fact, beyond the first solvation shell surrounding the ionic species, the solvent molecules behave like in the bulk solvent and do not contribute to $G_{\text{relax}}(t)$.

In summary, the present theoretical analysis predicts that the FIR absorption spectrum of an electrolytic solution is mainly governed by three mechanisms : (i) an induction mechanism modulated by the oscillatory motion of the solvated ions, (ii) a dipolar absorption mechanism due to the presence of ion pairs (CIP and SSIP) which vibrate and librate like diatomic molecules, (iii) a structural relaxation mechanism which expresses the reorganization of the solvent molecules during the solvation process.

In order to make a quantitative comparison with the experimental data, we have developed a band shape analysis based on the generalized Langevin equation. According to the Mori formalism (ref.8), the vibrational variables Q_γ , Q^{cip} , and Q^{SSIP} are considered as Brownian harmonic oscillators (anharmonicity is also taken into account). The oscillatory motion is characterized by a frequency mode, $\omega_\gamma = \sqrt{k_\gamma/m_\gamma}$, where k_γ is the force constant and m_γ the corresponding reduced mass, and by a damping function describing the translational diffusion of the species. In the case of solvated ions, k_γ is deduced numerically from the computer simulation data of Heinzinger et al. (ref.19) on alkali halide salts in aqueous solutions. In the case of ion pairs, Karim and Mc Cammon (ref.20) have investigated the hydration of a sodium chloride ion pair. By performing a Monte Carlo simulation, these authors have evaluated the solvent-averaged potential of mean force acting on each partner of the pair. Although these computer simulations only deal with aqueous solutions, the close similarity between water and methanol, as far as the solvation process is concerned, enable us to assume that the force constants k_γ , k_{cip} , and k_{SSIP} deduced from the aforementioned study give the correct order of magnitude for the alcoholic solution. For dipolar aprotic solvents (e.g. acetonitrile), the molecular dynamics simulation of Ciccotti et al. (ref.21), which evaluates the potential of mean force of an ion pair immersed in a solvent modelled by polar diatomic molecules, merits attention. This model system is loosely connected with real solutions and our use of the force constants, k_{cip} and k_{SSIP} , obtained by these authors must be understood as an attempt to test our theory. Finally, the determination of the absorption spectrum also requires the evaluation of several translational and rotational correlation functions (see eqns. 5-7). These latter ones can also be evaluated in the framework of the Mori formalism. The main effect of these

translational and rotational contributions is to broaden and to distort somewhat the band shape generated by the oscillatory modes.

DISCUSSION

The solvation bands corresponding to $\text{LiI}/\text{CH}_3\text{OH}$, $\text{LiI}/\text{CH}_3\text{CN}$, $\text{LiCl}/\text{CH}_3\text{OH}$ and $\text{RbI}/\text{CH}_3\text{OH}$, respectively, are presented in Fig.1. The absorption profile of lithium salts in methanol and acetonitrile solutions are characterized by three bands : a strong high frequency band that reaches the mid infrared range and two weak bands on the low frequency part of the spectrum. The high frequency peak (460 cm^{-1} in CH_3OH and 408 cm^{-1} in CH_3CN) is quite similar to the one observed by Maxey and Popov (refs.7,8) in their study on lithium salts in DMSO (where $\omega_{\text{max}} = 429\text{ cm}^{-1}$). This band was assigned to the vibration of the cation in a solvent cage. But these authors did not pay attention to the low frequency spectrum. Furthermore, our study shows that the low frequency bands are affected both, by the halide anion and by the solvent. When acetonitrile is substituted to methanol in LiI solutions, the Li^+ solvation band is shifted at lower frequency ($460 \rightarrow 408\text{ cm}^{-1}$), whereas the other two bands are strongly shifted in the opposite direction ($235 \rightarrow 362\text{ cm}^{-1}$, and $85 \rightarrow 145\text{ cm}^{-1}$, respectively). When chlorine is substituted to iodine in methanol solution, the Li^+ band and the intermediate band are mostly unaffected, but the band at 85 cm^{-1} shifts to 135 cm^{-1} . Finally, for RbI in CH_3OH , the absorption profile is mainly located at low frequency with a maximum around 85 cm^{-1} .

All the aforementioned spectral features can be understood in the framework of our theoretical analysis. The theory predicts four vibration bands corresponding to the oscillatory motions of the four ionic species (cation, anion, CIP, SSIP). Moreover, in the case of ion pairs (CIP and SSIP) each vibration band is convoluted with a rotation band which expresses the libration of the ion pair into a solvent cage. To illustrate our purpose, we show in Fig.2 the theoretical solvation band of $\text{LiI}/\text{CH}_3\text{CN}$. One notices (i) a high frequency peak at 410 cm^{-1} corresponding to the vibration of Li^+ , (ii) an intermediate peak at 340 cm^{-1} due to the stretching mode of CIP, (iii) a composite low frequency band ($0\text{--}120\text{ cm}^{-1}$) which is the superposition of the oscillatory mode of I^- (87 cm^{-1}), the stretching mode of SSIP (75 cm^{-1}), and the libration modes of CIP and SSIP (100 cm^{-1} and 35 cm^{-1} , respectively). The likeness between theoretical and experimental spectrum is striking. However, the relative contribution of each band to the calculated absorption intensity is only indicative since it is proportional to the concentration ratio of the ionic species, a quantity which is badly known. In the same way, the calculated bandwidths are related to the diffusion coefficients of the ionic species ; but, there also, these quantities are not accurately known. Nevertheless, the present theoretical analysis

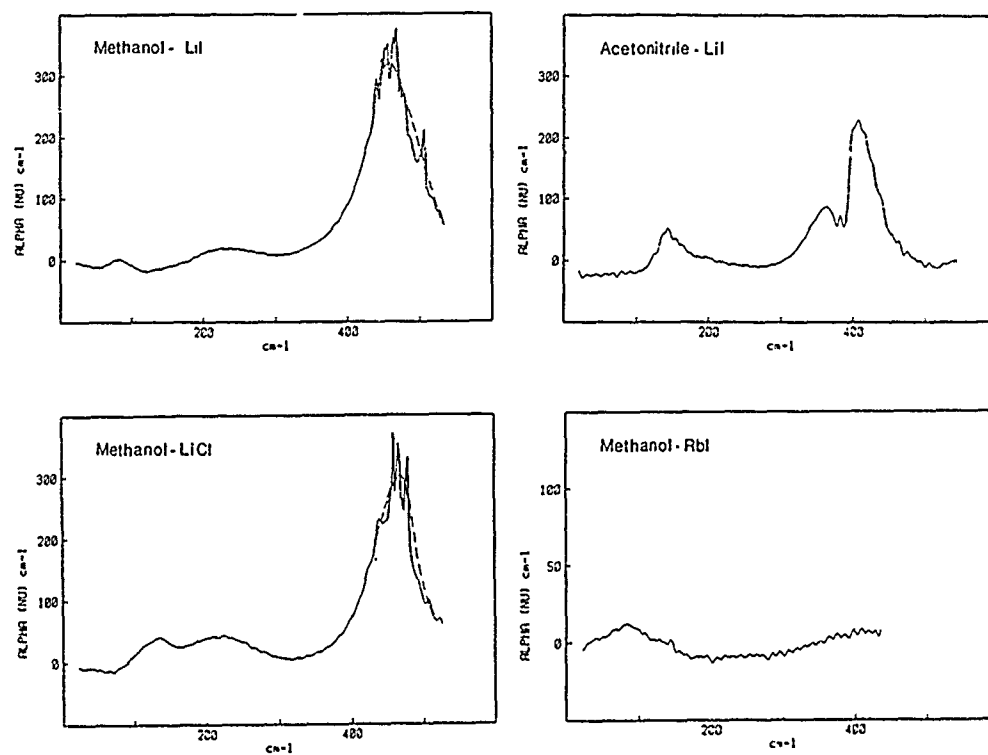
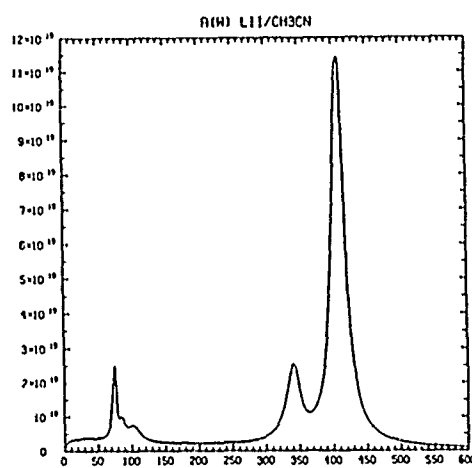


Fig. 1 Experimental absorption spectra.

Fig. 2 Theoretical absorption spectrum of $\text{LiI}/\text{CH}_3\text{CN}$.

elucidates the genesis of the spectrum, and quantitatively accounts for the solvent effect. For example, if one substitutes methanol to acetonitrile, the theory predicts that the Li^+ band is shifted at higher frequency ($410 \rightarrow 450 \text{ cm}^{-1}$) as a result of a tightening of the first solvation shell, while the stretching mode of CIP decreases ($340 \rightarrow 200 \text{ cm}^{-1}$) since the net force between the partners is weakened by the protic solvent. For RbI in methanol, the vibration bands collapse in a single band located below 120 cm^{-1} . All these predictions are confirmed by the experimental data.

REFERENCES

- 1 E. Grunwald, S. Highsmith and Ting-Po I, in "Ions and Ion pairs in Organic Reactions", vol.2, Ed. by M. Szwarc, Wiley (1974), p.447.
- 2 see all contributed papers in "Ions and Ion pairs in Organic Reactions", vol.1, Ed. by M. Szwarc, Wiley (1972).
- 3 G.W. Neilson, *Physica* 120B, p.325 (1983).
- 4 Proceedings of the NATO-ASI on "The Physics and Chemistry of Aqueous Ionic Solutions", Cargèse, Corsica (France, 1986), Ed. by M.C. Bellissent-Funel and G.W. Neilson, NATO ASI series vol.205, D. Reidel Publishing Company (1987).
- 5 J.C. Lestrade, J.P. Badiali and H. Cachet in "Dielectric and Related Molecular Processes", vol.2, A Specialist Periodical Report, p.106, The Chemical Society (London 1975).
- 6 W.F. Edgell, A.T. Watts, J. Lyford, IV, and W. Risen, Jr., *J. Am. Chem. Soc.*, **88**, 1815 (1966).
- 7 B.W. Maxey, A.I. Popov, *J. Am. Chem. Soc.*, **89**, 2230 (1967).
- 8 B.W. Maxey, A.I. Popov, *J. Am. Chem. Soc.*, **91**, 20 (1969).
- 9 J.L. Wuepper, A.I. Popov, *J. Am. Chem. Soc.*, **91**, 4352 (1969).
- 10 W.F. Edgell, J. Lyford, IV, R. Wright, W. Risen, Jr., and A. Watts, *J. Am. Chem. Soc.*, **92**, 2240 (1970).
- 11 A.T. Tsatsas, J.W. Reed, and W.M. Risen, Jr., *J. Chem. Phys.*, **55**, 3260 (1971).
- 12 T.L. Buxton, J.A. Caruso, *J. Phys. Chem.*, **77**, 1882 (1973).
- 13 R.G. Baum, A.I. Popov, *J. of Solution Chem.*, **4**, 441 (1975).
- 14 J.C. Evans, G.Y.S. Lo, *J. Phys. Chem.*, **69**, 3223 (1965).
- 15 J.M. Caillol, D. Levesque, and J.J. Weis, *J. Chem. Phys.*, **85**, 6645 (1986).
- 16 C.A. Krauss, *J. Phys. Chem.*, **60**, 129 (1956).
- 17 B. Guillot, Ph. Marteau, and J. Obriot, to be published
- 18 H. Mori, *Prog. Theor. Phys. (Kyoto)*, **33**, 423 (1965).
- 19 K. Heinzinger, *Physica*, **131B**, 196 (1985).
- 20 O.A. Karim, J.A. McCammon, *J. Am. Chem. Soc.*, **108**, 1762 (1986).
- 21 G. Ciccotti, M. Ferrario, J.T. Hynes, and R. Kapral, *Chem. Phys.*, **129**, 241 (1989).

DNA STRUCTURAL PARAMETER DETERMINATION BY VIBRATIONAL DATA

M. GHOMI

Laboratoire de Spectroscopie Biomoléculaire, U.F.R.
Biomédicale de Bobigny, Université Paris XIII, 93012 Bobigny
cédex (France)

SUMMARY

Normal coordinate analysis based on Raman and infrared data of oligo- and polynucleotides, allowed us to estimate the structural parameters (dihedral angles) of nucleosidic residues involved in DNA double helical chains. The present estimation derived essentially from the vibrational spectra recorded in solution (Raman) or in hydrated thin films (infrared) complete quite well those based on the other physical methods as NMR or X-ray diffraction.

INTRODUCTION

The purpose of this work is to show how the vibrational spectroscopy can be used as a probe in order to determine the molecular structural parameters. DNA has been selected as a characteristic example because of its high flexibility. The ability of this macromolecule to adopt right- and left-handed double helices, is now well known. Among the molecular structural parameters, torsion angles play the most important role as concerned with DNA conformational transitions. Change in these angles allows the double helix to go from one form to another. A conformational transition can be detected by vibrational spectroscopy by the change in the intensity and/or by the shift in the positions of Raman and infrared peaks.

From a classical viewpoint, the vibrating molecular system can be considered as a set of coupled oscillating point masses. Currently, a harmonic force field approximation has been used in order to express the potential energy of the vibrating molecule. This energy can be supposed to be invariant upon the conformational transitions. Thus, only the change in the kinetic energy, depending strongly on the

molecular geometry, is considered as responsible of the modifications observed in Raman or infrared spectra. Previously, a reliable simplified valence force field has been developed by our numerical investigations so as to reproduce the positions of the characteristic vibrational modes (vibrational markers) of the nucleosides (ref. 1-2), sugar (ref. 3) and phosphate-backbone (ref. 4).

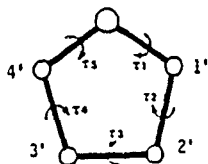
In the present work our aim is to demonstrate how the nucleosidic markers behave, upon the variation of the structural parameters. Our attention is especially focused on the manner how the structural parameters can be deduced from this kind of investigation.

NUCLEOSIDIC VIBRATIONAL MARKER ANALYSIS

Calculation details

Normal mode analysis is based on the Wilson GF-method (ref. 5). To perform the present calculations, an adequate code (NUCS) has been elaborated. Redundancy among the internal coordinates has been entirely resolved. Numerical computations have been carried out both on a DPX-network (BULL) and a CRAY-2 computer. The CRAY version has been vectorized and multitasked in order to improve its run-time. It turns about 340 times faster than the scalar DPX version.

Taking account of their important contribution to the DNA vibrational markers, the purine deoxynucleotides, i.e. dG and dA, are studied in the present investigation. For both of these residues, dynamic models as shown on figure 1 have been used. They are constituted by the guanine and adenine residues associated to 2'-deoxyribose sugars. The dynamic models are extended up to their O3' and O5' terminals. The orientation of the base is determined by the glycosidic torsion angles, i.e. χ (O1'-C1'-N9-C4). Two distinct conformations can be considered, namely anti ($180^\circ < \chi < 300^\circ$) and syn ($0^\circ < \chi < 90^\circ$). The sugar conformation depends on the five dihedral angles around the ribose ring bonds, which can be estimated by the following expression:



$$\tau_i = \tau_m \cos (P + (i+2) 4\pi/5)$$

$$i = 1, 2, 3, 4, 5$$

where τ_m represents the puckering amplitude (deviation from a planar sugar) and P denotes the phase angle of pseudorotation characterizing the ring conformation. Two types of sugars are recognized, namely N(north)-type ($P=0^\circ\pm 90^\circ$) and S(south)-type ($P=180^\circ\pm 90^\circ$) (fig. 2.).

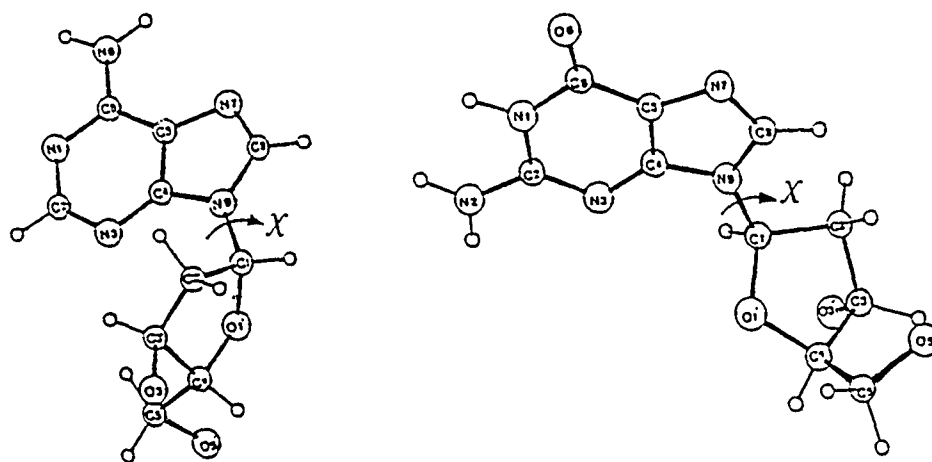


Fig. 1. The dA residue (left) as involved in Z DNA (C3'-endo/syn conformation) and the dG residue as encountered in B DNA (C2'-endo/anti conformation).

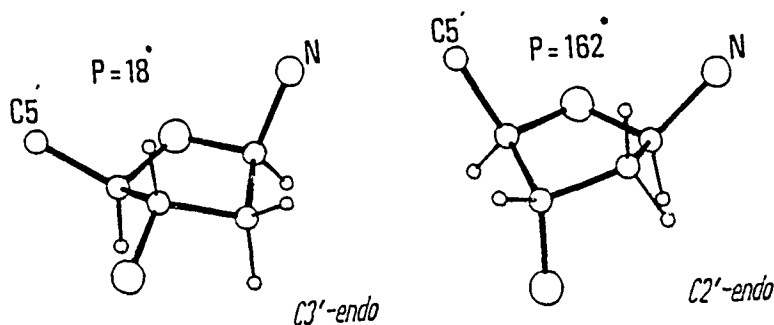


Fig. 2. Influence of the pseudorotation phase angle on the sugar pucker conformation.

Normal mode analysis has been performed by successive jumps of 36° (for P angle) and 30° (χ angle), respectively. τ_m angle has been fixed to 40° , which represents a characteristic value for the sugars found in DNA (ref. 3).

Marker modes and their behaviour versus structural parameters

Four of the dA and dG vibration markers have been selected. They have been detected in the quasi-whole oligo- and polynucleotide vibrational spectra and give rise to the well resolved infrared and/or Raman bands (ref. 6-10). These modes have been numbered from 1 to 4. And their positions as detected in A, B and Z forms of poly d(G-C) and poly d(A-T) have been mentioned in table 1. The most striking effect is that these modes behave in exactly the same manner for the dA and dG residues upon the DNA conformational transitions. As it is shown in table 1, the mode 1 is Raman active, while the mode 2 is observed in infrared spectra. The mode 3 can be detected in both Raman and IR spectra. Its evolution is however more clear in IR spectra and this is the reason why it is known rather as an infrared marker. Finally, the mode 4 concerns the most discussed Raman marker taking account of its important shift upon the B to Z conformational transition.

TABLE 1

Experimental positions (cm^{-1}) of the four marker modes studied in the present work.

Markers	Polynucleotides				
	poly d(G-C)		poly d(A-T)		
	B	Z	A	B	Z
mode 1	1422 ^a	1430 ^a	1415 ^b	1419 ^c	1440 ^c
mode 2	1420 ^d	1409 ^d	1416 ^e	1425 ^e	1408 ^e
mode 3	1374 ^d	1354 ^d	1374 ^e	1374 ^e	1357 ^e
mode 4	681 ^a	625 ^a	662 ^b	666 ^c	622 ^c

^aref. 6

^bref. 9

^cref. 10

^dref. 7

^eref. 8

The three dimensional representation of the selected marker frequencies obtained by our normal coordinate analysis for the dG residue have been shown in figure 3. The P parameter does not considerably affect the mode 1. While both P and χ parameters affect the mode 2. The mode modes 3 and 4 are also dependent of these conformational parameters. A striking shape of variation has been obtained for mode 4. A very deep valley is found in the region of the N-type sugar connected to the syn bases (Z form). A less deep minimum corresponding to the low anti bases (A form) appears on the surface relative to the mode 4.

Similar surfaces have been obtained for the dA residue marker modes. An extended report of these calculations accompanied by more detailed information concerning the normal mode assignments will be published in a forthcoming publication.

Estimation of the nucleosidic structural parameters

Obviously, the structural parameters of the nucleosides involved in the right- and left-handed double helices can be estimated by comparing the experimental (table 1) and calculated (fig. 3) results. For a given helical conformation, P and χ parameters should be extracted from the regions of the surfaces which give calculated wavenumbers satisfying the experimental criteria for all of the four markers. The results of this estimation is presented in tables 2 and 3. This structural determination based on the vibrational spectroscopy is also in good agreement with that derived from X-ray diffraction or NMR spectra.

In conclusion, the present method allows to make a quantitative determination of the structural parameters related to the oligo- and polynucleotides found in solid, fibrous or liquid phases. Moreover, the study of the evolution of the vibrational modes versus the dA residue conformations led us to predict the position of its marker modes in Z form (ref. 11). The so-obtained calculated results could be verified by the Raman spectrum of poly d(A-T) adopting Z helix conformation (ref. 10).

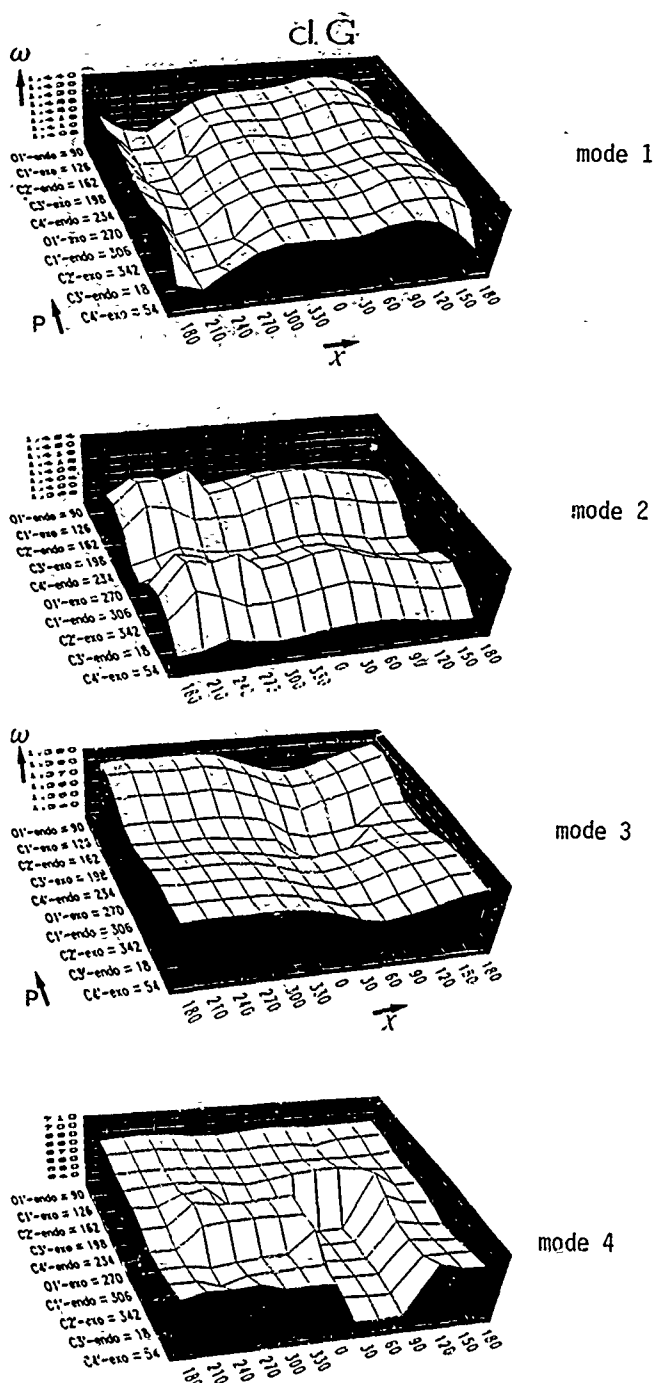


Fig. 3 Three dimensional surfaces representing the variation of the dG residue marker mode wavenumbers (cm^{-1}) as a function of the P and χ (degrees) structural parameters.

TABLE 2

Estimation of the P and X parameters (in degrees) for the dG residues as involved in different conformations of poly d(G-C).

DNA conformation	P	X	Nucleosidic conformation
B	162±18	270±15	C2'-endo/high anti
Z	18±36	60±15	C3'-endo/syn

TABLE 3

Same as table2 but for the dA residue involved in different conformations of poly d(A-T).

DNA conformation	P	X	Nucleosidic conformation
A	18±36	210±30	C3'-endo/low anti
B	162±36	270±15	C2'-endo/high anti
Z	18±36	60±15	C3'-endo/syn

REFERENCES

- 1 R. Letellier, M. Ghomi and E. Taillandier, Interpretation of DNA vibration modes: I-The guanine and cytidine residues in poly d(G-C) and d(CG)₃, J. Biomol. Str. Dyn. 3(4) (1986) 671-687.
- 2 R. Letellier, M. Ghomi and E. Taillandier, Interpretation of DNA vibration modes: II- The adenine and thymidine residues involved in oligo- and polynucleotides, J. Biomol. Str. Dyn. 4(4) (1987) 663-683.

- 3 D. Dohy, M. Ghomi and E. Taillandier, Interpretation of DNA vibration modes: III- The behaviour of the sugar pucker vibration modes as a function of its pseudorotation parameters, *J. Biomol. Str. Dyn.* 6(4) (1989) 741-754.
- 4 R. Letellier, M. Ghomi and E. Taillandier, Interpretation of DNA vibration modes: IV- A single helical approach to assign the phosphate-backbone contribution to the vibrational spectra in A and B conformations, *J. Biomol. Str. Dyn.* 6(4) (1989) 755-768.
- 5 E.B. Wilson Jr., J.C. Decius and P.C. Cross, *Molecular Vibrations*, Dover, New York, 1980.
- 6 T.J. Thamann, R.C. Lord, A.H.J. Wang and A. Rich, The high-salt form of poly d(G-C) in left handed Z DNA: Raman spectra of crystals and solutions, *Nucl. Acids Res.* 9(20) (1981) 5443-5457.
- 7 J.A. Taboury, J. Liquier and E. Taillandier, Characterization of DNA structures by infrared spectroscopy: double helical forms of poly d(G-C), poly d(D₈G-C) and poly d(G-m⁵C), 63 (1985) 1904-1909.
- 8 S. Adam, J. Liquier, J.A. Taboury and E. Taillandier, Right- and Left- handed helices of poly d(A-T) investigated by infrared spectroscopy, *Biochemistry* 25 (1986) 3220-3225.
- 9 G.J. Thomas Jr. and J.M. Benevides, An A-helix structure for poly d(A-T), *Biopolymers* 24 (1985) 1101-1105.
- 10 J.P. Ridoux, J. Liquier and E. Taillandier, Raman spectroscopy of Z-form poly d(A-T), *Biochemistry* 27 (1988) 3874-3878.
- 11 M. Ghomi, R. Letellier and E. Taillandier, A critical review of the nucleosidic vibration modes appearing in the 800-500 cm⁻¹ spectral region based on new harmonic dynamics calculations, *Biopolymers* 27(1988) 605-616.

FORCE FIELD CALCULATIONS OF SIDEROPHORES.
STABILITY AND CONFORMATIONS OF FeIII CHELATES WITH
NOVEL IRON RELATED PARAMETERS.

A. BOURAQUI¹, M. FATHALLAH¹, F. M'HENNI², B. BLAIVE¹,
AND R. GALLO¹.

¹ ESIPSOI, UA126CNRS, Fac Sciences St Jérôme, 13397 Marseille
cedex 13, France.

² Fac Sciences and Tech, Dept Chemistry, route de Kairouan,
Monastir Tunisia.

SUMMARY

Quantitative estimations of conformations and energies of iron carriers (siderophores) are made by molecular mechanics. New iron related parameters are designed to be used with the 1985 Allinger MM2 parametrization. They are tested with new FeIII chelates containing catechol and carboxylic acid subunits. The results help in tailoring new ligands and encourage further force field calculations of iron and other high valent metal chelates.

INTRODUCTION

Iron is essential for life. In human a lack of the element causes anemia but an excess is toxic (ref 1). On the other hand in plants a lack of the element causes ferric chlorosis (ref 2). Iron carriers (siderophores) are organic ligands able to complex and transport FeIII for establishing an appropriate level of the element (ref 3). So far no quantitative study relating the structure of these molecules to their energy of association with the ferric ion has been carried out. It is the purpose of this paper to describe the new parameters designed for, and the calculations made with, the molecular mechanics procedure used to study this question.

MOLECULAR MECHANICS CALCULATIONS

The method

Molecular Mechanics methods (also called Empirical Force Field methods) have been used increasingly to study molecular structure

and properties (refs 4-7). These methods are based on the assumption that the energy of the molecules can be expressed by calculations of empirical nature based on classical mechanical principles. The total energy of the molecular system under study is treated as the sum of several components

$$E(\text{total}) = E(l) + E(\vartheta) + E(\phi) + E_{nb}(r)$$

where $E(l)$ = bond length deformation, $E(\vartheta)$ = bond angle deformation ("Bayer strain"), $E(\phi)$ = torsional eclipsing energy ("pitzer strain") and E_{nb} = non bonded or Van der Waals interactions. Adjustable parameters have been evaluated to treat all energy components. In the 1985 Allinger MM2 program (ref 8), presently the most widely used, parameters are available for carbon, hydrogen, and most of the heteroatoms but not yet for Iron. We have first designed new parameters for treating the ferric ion.

Design of parameters related to FeIII.

Since the complexation of iron by organic ligands is made through oxygen atoms E_l and E_ϑ involve only FeIII-oxygen interactions.

(i) Iron-oxygen bond length elongation energy E_l . It is calculated by Hooke's law $E(l) = 1/2 \sum k_l (l - l_0)^2$. The bond length Fe-O is taken from X-ray data (ref 9) as $l_0 = 1.9955 \text{ \AA}$. The constant k_l is estimated from IR data of $[\text{FeF}_6]^{3-}$ where $\bar{\nu} = 538 \text{ cm}^{-1}$ relates to Fe-F stretching (ref10); since $\bar{\nu} = 1/2\pi \sqrt{k/\mu}$ and $\mu = 1/m_1 + 1/m_2$ (where μ = reduced mass, m_1 mass of iron, m_2 mass of fluor) then $k = 4\pi^2 \nu^2 c^2 = 3.260 \text{ mdyne/\AA}$.

(ii) Bending (bond angle deformation) energy E_ϑ . Also calculated by Hooke's law $E(\vartheta) = 1/2 \sum k_b (\vartheta - \vartheta_0)^2$. The constant k_b also derived from IR data of $[\text{FeF}_6]^{3-}$ (ref10) gives $k_b = 2.039 \text{ mdyne \AA/rd}^2$ from $\bar{\nu} = 248 \text{ cm}^{-1}$. ϑ_0 is derived from an octahedral structure of the iron complex where the angle O-Fe-O = 90°

(iii) Torsional eclipsing energy E_ϕ . Estimated by an equation of the form

$$E(\phi) = 1/2 \sum (v_1 (1 + \cos\phi) + v_2 (1 - \cos 2\phi) + v_3 (1 + \cos 3\phi) + \dots)$$

v_1 , v_2 and v_3 were obtained by making the following approximations:

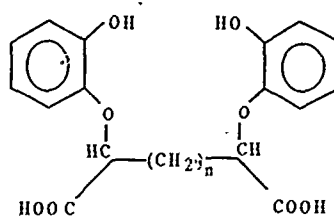
Atoms involved in torsional eclipsing energy	Parameters used			Taken from "equivalent" series	In reference
	v1	v2	v3		
1-1-6-26	.8	0	.09	1-1-6-5	ref 11
3-1-6-26	0	0	.09	3-1-6-5	ref 11
5-1-6-26	0	0	.2	5-1-6-5	ref 11
2-2-6-26	1	1.65	0	2-2-6-5	ref 11
1-3-6-26	0	.5	0	1-3-6-5	ref 11
7-3-6-26	-3.28	5.6	0	7-3-6-5	ref 11
1-6-26-6	0	0	.33	2-2-25-2	ref 12
2-6-26-6	0	0	.33	2-2-25-2	ref 12
3-6-26-6	0	0	.33	2-2-25-2	ref 12

The atoms code is: 1= C_{sp}³, 2= C_{sp}², 3= C(of CO), 5= H, 6= doubly bonded oxygen, 7= O (of CO), 25= P, 26= Fe.

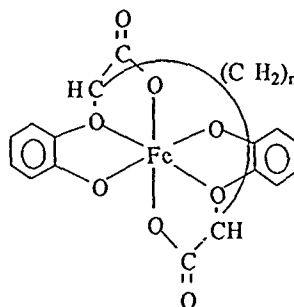
(iii) Non bonded or Van der Waals energy E_r. Based on a "Hill function" $E(r) = \epsilon (- C_1 (r^*/r^6) + C_2 \exp (- C_3 (r/r^*)))$ where ϵ is an estimate of iron "hardness"; r^* the sum of the Van der Waals radii of interacting atoms; r the interatomic distance; C_1 , C_2 , and C_3 numerical constants. In the MM2 parametrization (ref 8), ϵ is usually estimated empirically. Iron being a bulky atom we have estimated a value $\epsilon = .17$ close to the value used for Phosphorus. For evaluating r^* we have firstly plotted V.d.W. radii of several elements as a function of atomic numbers and observed parallel lines for transition elements; the extrapolation of the line corresponding to the fourth period gives $r^*(Fe) = 1.82 \text{ \AA}$, from which for FeIII, if one refers to vanadium (v), we have obtained 2.2 \AA .

However V.d.W. radii used by Allinger in MM2 are systematically different of usual V.d.W. values proposed by Bondi. In fact a close examination of the correlation between MM2 values by Allinger versus "physical" values by bondi reveals a very good linear regression of slope 0.856 with a correlation coefficient $R=0.994$. From this correlation the FeIII radius suitable for MM2 calculations is estimated to be $r^*(Fe^{3+}) = 2.391 \text{ \AA}$.

(iiii) Treatment of the iron-oxygen interaction. The series of iron complexes that we have studied are hexadentates. The atoms directly bonded to iron are oxygen atoms. We have termed the ligands AP_n (A for Acid, P for pyrocatechol, n for the number of methylen units) (ref 14).



The interaction of iron with these ligands gives a complex [Fe^{III} ligand] which has an octahedral structure.



scheme 1

In this molecular arrangement two different iron-oxygen interactions have to be considered:

-When the oxygen atom is bonded to only one atom other than iron it bears a negative charge : $\text{C}-\text{O}^-\cdots\text{Fe}^{3+}$

In this case the curve $E = f(l)$ corresponds to a parabola $E = k(1 - l_0)^2$ where $k = 3.26 \text{ m dyn} \cdot \text{\AA}^{-1} = 4.67 \cdot 10^{-2} \text{ kcal} (\text{\AA}^2)^{-1} \cdot \text{mole}$ and where $l_0 = 1.9955 \text{\AA}$ which is a mean value of Fe-O distances from X-ray data.

-When the oxygen atom is bonded to two atoms others than iron

it bears no charge : $\text{C}=\text{O} \cdots \text{Fe}^{3+}$ and in this case the curve $E = f(l)$ may be derived from a minimum in energy of $-3 \text{ kcal} \cdot \text{mole}^{-1}$ (often observed in iron heteroatom dative bonds) at a distance Fe-O of 1.9955\AA , and from a value $E=0$ at distances 1.977 and 2.014\AA corresponding to extremes values of Fe-O bonds in X-ray structures (ref 15). The equation of the parabola then becomes

$$E = -3 + k(1 - 1.9955)^2$$

with $k = 0.83 \cdot 10^4 \text{ kcal} (\text{\AA}^2)^{-1} \cdot \text{mole}^{-1} = 58 \text{ m dyn} \cdot \text{\AA}^{-1}$.

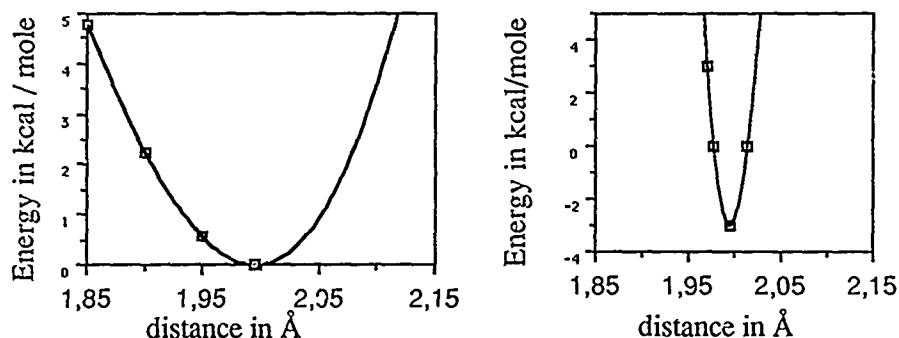
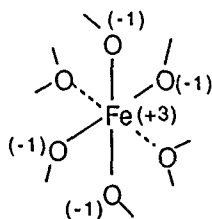


Fig.2 $E=f(l)$ for "negative" oxygens Fig.3 $E=f(l)$ for "neutral" oxygens

The steep shape of the curve means consequently that the "neutral" (doubly bonded) oxygens will be kept close to the iron atom, (and that the V.d.W. interaction between these two atoms are omitted in the estimation of the total energy).

Treatment of the electrostatic forces.

Molecular mechanics methods have been applied recently for estimating complexation constants of alkaline, alkaline earth and other metallic cations with organic ligands (refs 16-18). The complexation of iron III was not treated except in a recent paper by Lifson *et al* who used a Force Field Method where they incorporated the electrostatic forces (ref 19). Taking the case of our series of complexes to exemplify the procedure:



The total energy charge of ion is : $+3+4(-1)+2(0)=-1$ (however a fractionnal negative charge ≈ -0.2 (ref 20) could be given to "neutral" oxygens to represent the electrostatic attraction between iron and oxygen non bonding pairs). The iron atom facing the electron pairs of "neutral" and "negative" oxygen, the same V.d.W. repulsive potential may be used in both cases; a (r^{-6}, exp) or a (r^{-9}, r^{-6}) expression. The "Hill function" in (r^{-6}, exp) used by Allinger, incorporating the V.d.W. radius and the "hardness" of iron, gives a strong destabilizing

energy at lengths near 2\AA close to those of Fe-O bonds. The (r^{-9}, r^{-6}) potential ($V=A/r^9 - C/r^6$) may also be used for the modelisation of the electrostatic potential (ref 22); in this case the A and C coefficients are deduced from iron and oxygen parameters by two calculation procedures giving almost the same result.

The potential $V_{9-6} = \epsilon (2(r^*/r)^9 - 3(r^* - r)^6)$

with iron parameters : $\epsilon = .0192$; $r_1 = 3.84\text{\AA}$; charge = +3

and oxygen parameters : $\epsilon = .0198$; $r_2 = 3.65\text{\AA}$; charge = -1

gives the curve of fig 4

whereas the potential $V_{9-6} = A/r^9 - C/r^6$

with iron coefficients $A_1 = 2\epsilon r^{*9} = 6.9710^3$; $C_1 = 3\epsilon r^{*6} = 1.85 \cdot 10^2$

with oxygen coefficients $A_2 = 2\epsilon r^{*9} = 4.55 \cdot 10^4$; $C_2 = 3\epsilon r^{*6} = 1.40 \cdot 10^3$ and with iron-oxygen coefficient $A = \sqrt{A_1 A_2}$; $C = \sqrt{C_1 C_2}$

gives the curve of fig 5

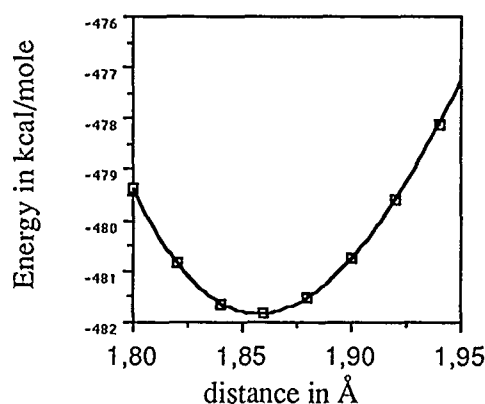


fig 4. $E=f(l)$ using ϵ and r^*

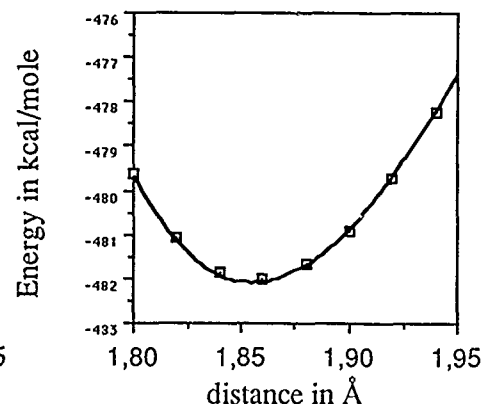


fig 5. $E=f(l)$ using A and C

The comparison of potential energy curves 1, 3 and 4 shows a different value of the minimum in energy but a very similar shape of the curve near the value l_0 . In other words the geometries calculated by two methods would be very similar but the absolute values of the strain energy would differ notably. In addition the value $l_0 = 1.87\text{\AA}$ used by Lifson seems short compared to known X-ray Fe-O bond lengths; but on the other hand the shape of the figure 3, that we used, is sharp, which means a narrow range of variation in bond length.

In short the procedure described in this paper and that reported by Lifson should give very similar geometries, slightly different absolute values but very comparable relative values in strain energy. When the purpose is to compare relative complexing abilities in given series both procedures should be helpful.

RESULTS AND DISCUSSION

We have carried out the calculations on HP 9000-825 computer with BLEMO (ref 21-22), an adapted molecular mechanics program, written in FORTRAN 77, containing 3731 lines and accepting molecules with 135 atoms. The parameters related to iron are those described in this paper, otherwise we have used the 1985 MM2 parameters described by Allinger *et al* (ref 8). The minimization used a step by step relaxation procedure carried out on internal coordinates.

The purpose of the calculations was to find the most efficient iron complexant in the series AP_n. These molecules, which have some "similarity" with EDTA derivatives, contain two carboxy- α -catechol moieties separated by a spacer of several methylene units; The energy of the optimized geometry of iron chelates was obtained by starting from the fully symmetrical octahedral structure (scheme 1) built up with "natural" geometrical parameters, and by reaching a minimum in energy after full relaxation. The figure 6 shows clearly that the most stable iron chelate is obtained when the number of methylene units is $n=4$

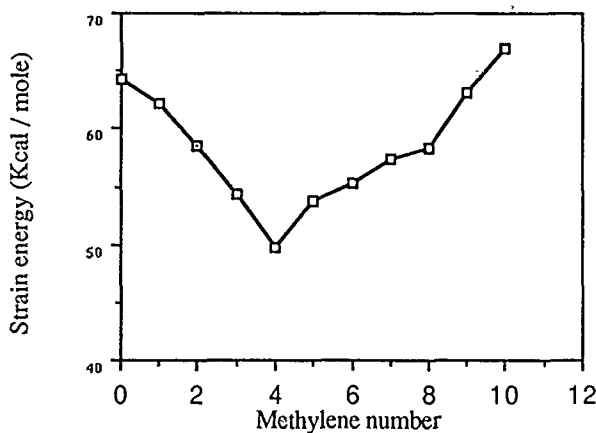


fig.6 Energy vs n in AP_n

An examination of the possible isomers existing in each case shows 4 geometrical positions depending upon the location of the catechol and carboxylic function in equatorial or axial position (ref 23). Moreover each geometrical isomer exists in threo or erythro configuration. In the case of AP₄ the complex of lowest energy corresponds to the threo form; we have therefore undertaken the

preparation of the appropriate isomers in the series APn by a synthetic route involving α -bromination of an α - ω -dicarboxylic acid, followed by nucleophilic substitution with gallicol under phase transfer catalysis conditions and deprotection of the methyl ether; the syntheses are in course (ref 14)

In addition several series of siderophores which have structures homomorphic to enterobactin, like CYCAM (ref 24), TRENCAM (ref 25), or MECAM (ref 26), have been prepared. All of them contain 3 catechol units attached to a spacer of different nature. We are using the calculations described in this paper to design and prepare new iron complexants of this type.

CONCLUSION

In summary we have extended molecular mechanics calculations to the series of iron complexants APn using the BLEMO computer program with MM2 parameters and novel iron related parameters. Calculated geometries are consistent with previous X-ray determinations of similar derivatives. The relative energies allow to select the most efficient complexant in the series. The method may be helpful in tailoring new ligands.

REFERENCES

- (1) W. H. Crosby, "Iron absorption" in Best and Taylor's "Physiological Basis of Medical Practice", J. R. Brobeck Ed, Williams and Wilkins, Baltimore, 1979.
- (2) A. Wallace, Journal of plant nutrition, 1982, 5, 277.
- (3)a R. C. Hider, Struct. and Bonding, 1984, 58, 25
- (3)b K. N. Raymond and T. M. Garrett, Pure & Appl Chem, 1988, 60, 1807
- (4) U. Burkert, and N. L. Allinger, "Molecular Mechanics", ACS series, Washington D.C., 1982
- (5) E. Osawa and H. Musso, Angew. Chem. Int. Ed. Engl, 1983, 22, 1.
- (6) E. Osawa and H. Musso, Top. Stereochem., 1982, 13, 117.
- (7) J. S. Lomas, L'actualité chimique, 1986, 5, 7.
- (8) N. L. Allinger, Molecular Mechanics MM2 program Q. C. P. E., 1985
- (9) R. C. Scarrow, P. E. Riley, K. Abu-Dari, D. L. White and K. N. Raymond, Inorg. Chem., 1985, 24, 954.
- (10) K. Kakamoto, "Infrared and raman spectra of inorganic and coordination compounds 3th Ed, Wiley Interscience, N. Y., 1978.
- (11) N. L. Allinger and Y. H. Yuth, Molecular Mechanics Program Q.C.P.E. 395, 1980
- (12) J. Phillip-Bowen and N. L. Allinger, J. Org. Chem. 1987, 52, 2937.
- (13) A. Bondi, J. Phys. Chem., 1964, 68, 441.
- (14) F. Mhenni, A. Bouraoui, Z. Mighri and R. Gallo, Bull. Soc. Chim. Fr., in press.
- (15) T. J. Mc Murry, M. W. Hosseini, T. M. Garrett, F. E. Hahn, Z. E. Reyes and K. N. Raymond, J. Am. Chem. Soc. 1987, 109, 7196.
- (16) S. Lifson, C. E. Felder and A. Shanzer, J. Am. Chem. Soc, 1983, 105, 3
- (17) T. W. Hambley, G. H. Searle; Acta crystallogr., 1984, 78, 75
- (18)a R. D. Hancock, J. G. Mc Dougall, F. Marsicano; Inorg. Chem., 1979, 18, 10.
- (18)b P. D. J. Grootenhuis, P. A. Kollman; J. Am. Chem. Soc., 1989, 111, 2152.
- (18)c G. Wipff, P. Weiner and P. A. Kollman; J. Am. Chem. Soc., 1982, 104, 3249.
- (18)d P. A. Kollman, G. Wipff and C. Singh; J. Am. Chem. Soc., 1985, 107, 2212.

- (19) A. Shanzer, J. Libman, S. Lifson, C. E. Felder, J. Am. Chem. Soc., 1986, **108**, 7609.
- (20) G. Pèpe, B. Serres, D. Laporte, G. Del Re, C. Minichino, J. Theo. Biol.; 1985, **115**, 571
- (21) C. Roussel, B. Blaive, R. Gallo, J. Metzger, J. Sandström, Org. Magn. Reson., 1980, **14**, 166.
- (22) Available on request to Bruno Blaive.
- (23) F. Mhenni, A. Bouraoui, M. Fathallah, B. Blaive and R. Gallo; J. Chem. Soc. Perk 2; in press.
- (24) W. R. Harris and K. N. Raymond, J. Am. Chem. Soc, 1979, **101**, 2728.
- (25) S. J. Rodgers, C. W. Lee, C. Y. Ng and K. N. Raymond, Inorg. Chem, 1987, **26**, 1622.
- (26) F. L. Weill and K. N. Raymond, J. Am. Chem. Soc, 1979, **101**, 6534.

DISCUSSION

DEVILLERS - Is it possible to use in the program BIG STRAIN 3 the parameters designed for iron in this study ?

BOURAOUI - BIG STRAIN " is one of the major Molecular Mechanic (MM) programs available on QCPE ; but each program runs with its own parametrization. The parameters designed in this study are fitted to the Allinger MM2 parameters. However the same methodology may be followed to design iron parameters suitable for an other set of constants.

DEVILLERS - Is the octahedral structure of the 6 oxygens surrounding the iron atom symmetrical or not ?

BOURAOUI - The calculation is started with an octahedral structure fully symmetrical, but when a minimum in energy is reached after relaxation the octahedron is slightly distorted. This result is in agreement with recent X-ray studies by Raymond et al on other siderophores.

BRUNEAU - How is made exactly the parametrization of iron ?

BOURAOUI - From geometrical and spectroscopic data : the V d W radius of iron is derived from a number given by Bondi, corrected to fit the other MM2 values. the reference length l_0 is derived from X ray data. The constants k_1 and k_b are derived from IR data.

LAMBERT - What is the interest of these ligands ?

BOURAOUI - The goal is to control the concentration level of iron by complexing and carrying this element both in humans and in plants.

DEMANGE - Is it possible to predict the energy of other iron complexes with ligands having different structures and to compare these values ?

BOURAOUI - The calculations that we have carried out are related to the same molecular series, and therefore the relative values that we obtain are reasonably accurate. Absolute values of energy obtained with totally different structures would be very useful but need at the moment further comparison with experimental data.

DEMANGE - Could you treat other ligands where the ferric ion is complexed through sulfur atoms ?

BOURAOUI - In principle there is no fundamental objection and no major difficulty for designing the appropriate parameters and for doing these calculations.

MEHANI - What industrial applications do you expect for these iron complexants ?

BOURAOUI - They may be helpful in agriculture for treating iron chlorosis, a lack of iron in plants that turns the leaves yellow, and which further makes the plants dry. In addition they may be helpful in humans for increasing the level of iron, become low from frequent blood transfusion, which is toxic.

MEHANI - Is the calculation method easy to operate ?

BOURAOUI - Yes, the program has been devised to be run by the non specialist. It is written in FORTRAN 77 ; it contains 3731 lines and accepts 135 atoms. It is conversational, it can be modified easily and it optimizes more rapidly than the popular 1985 Allinger MM2 program.

FOURCOT - Is it possible to extend the calculations to other alkaline, alkaline-earth or metallic cations ?

BOURAOUI - Using the procedure that we have followed this extension is certainly possible, provided the appropriate parameters can be designed.

FOURCOT - What is exactly the MM method that you have used ?

BOURAOUI - It is named BLEMO ; it makes the usual minimization in energy of the classical mechanical terms including stretching, bending, eclipsing and Van der Waals. The parameters for iron are described in this study ; the parameters for C,H,O are from Allinger MM2. The program is available on request to B. Blaive.

GRAND - Where do the molecules that you have studied come from ?

BOURAOUI - We have made the preparation of these molecules and described the synthetic methods, which need the protection-deprotection of carboxylic acid and guaiacol, in a Bull Soc Chim paper to be published soon.

GRAND - Did you compare your calculated geometries with X-ray values ?

BOURAOUI - We plan to do that as soon as we will have monocrystals of sufficient size. Presently the crystals that we obtain are too small for an X-ray analysis.

WARSHEL - How does the experimental complexation constants compare to the calculated energies of the complexes ?

BOURAOUI - We are presently meeting difficulties with these measurements because of the low solubility of the complexes in water ; the problem is being solved by increasing the hydrophilicity of the complexes via sulfonation of the aromatic rings.

WIPFF - Do you consider the charges on the oxygens and on the iron ?

BOURAQUI - The calculations consider only the Van der Waals interactions. However we have compared two procedures : the first without the charges and the second with the charges. The two approaches give comparable potential curves and consequently similar relative energies.

RELATION ENTRE CONSTANTES DE FORCE ET CONSTANTES ELASTIQUES DANS DIVERS MATERIAUX

E. HUSSON, Y. REPELIN et D. WEIGEL

Laboratoire de Chimie Physique du Solide, URA CNRS D0969, Ecole Centrale de Paris, 92295 Chatenay-Malabry Cedex (France)

SUMMARY

A general matrix method is used for calculating vibrational frequencies and elastic constants from a generalized valence force field (GVFF). The validity of the method is discussed and the most important parameters influencing the elastic behaviour of the studied compounds are pointed out.

INTRODUCTION

Le but de cette étude est de relier l'élasticité macroscopique d'un composé, définie par les éléments C_{ij} du tenseur d'élasticité, à l'élasticité microscopique, définie par les constantes de force d'élongation de liaisons, de déformations angulaires, de torsion... d'un champ de force de valence généralisé.

METHODES UTILISEES

La détermination des constantes de force à partir des spectres de vibration expérimentaux est basée sur la méthode des matrices GF de Wilson (ref. 1) étendue au cristal selon la théorie de Shimanouchi et coll (ref. 2).

Les éléments C_{ij} du tenseur d'élasticité ont été calculés à partir des constantes de force par la méthode matricielle développée par Shiro et coll (ref. 3) à partir de la théorie de Born et Huang (ref. 4). La matrice du tenseur d'élasticité est donnée par la formule générale :

$$C = \{ \tilde{D}_\sigma F_R^\sigma D_\sigma - \tilde{D}_\sigma F_R^\sigma B_p [\tilde{B}_p F_R^\sigma B_p]^{-1} \tilde{B}_p F_R^\sigma D_\sigma \} / v$$

dans laquelle F_R^σ est la matrice de l'énergie potentielle (constantes de force), D_σ est la matrice dynamique définie à partir de la matrice des coordonnées internes et des positions atomiques; B_p est la matrice de transformation des coordonnées cartésiennes en coordonnées internes et v est le volume de la maille cristalline. Le programme que nous avons écrit calcule de plus les dérivées partielles des constantes élastiques par rapport aux constantes de force (distribution de l'énergie potentielle ou DEP).

Les calculs ont été effectués sur un ordinateur UNIVAC 1110.

Le champ de force peut ainsi être déterminé à partir des fréquences expérimentales de vibration ou (et) à partir des éléments C_{ij} expérimentaux. Dans cette étude, nous nous sommes intéressés à des composés dont les valeurs expé-

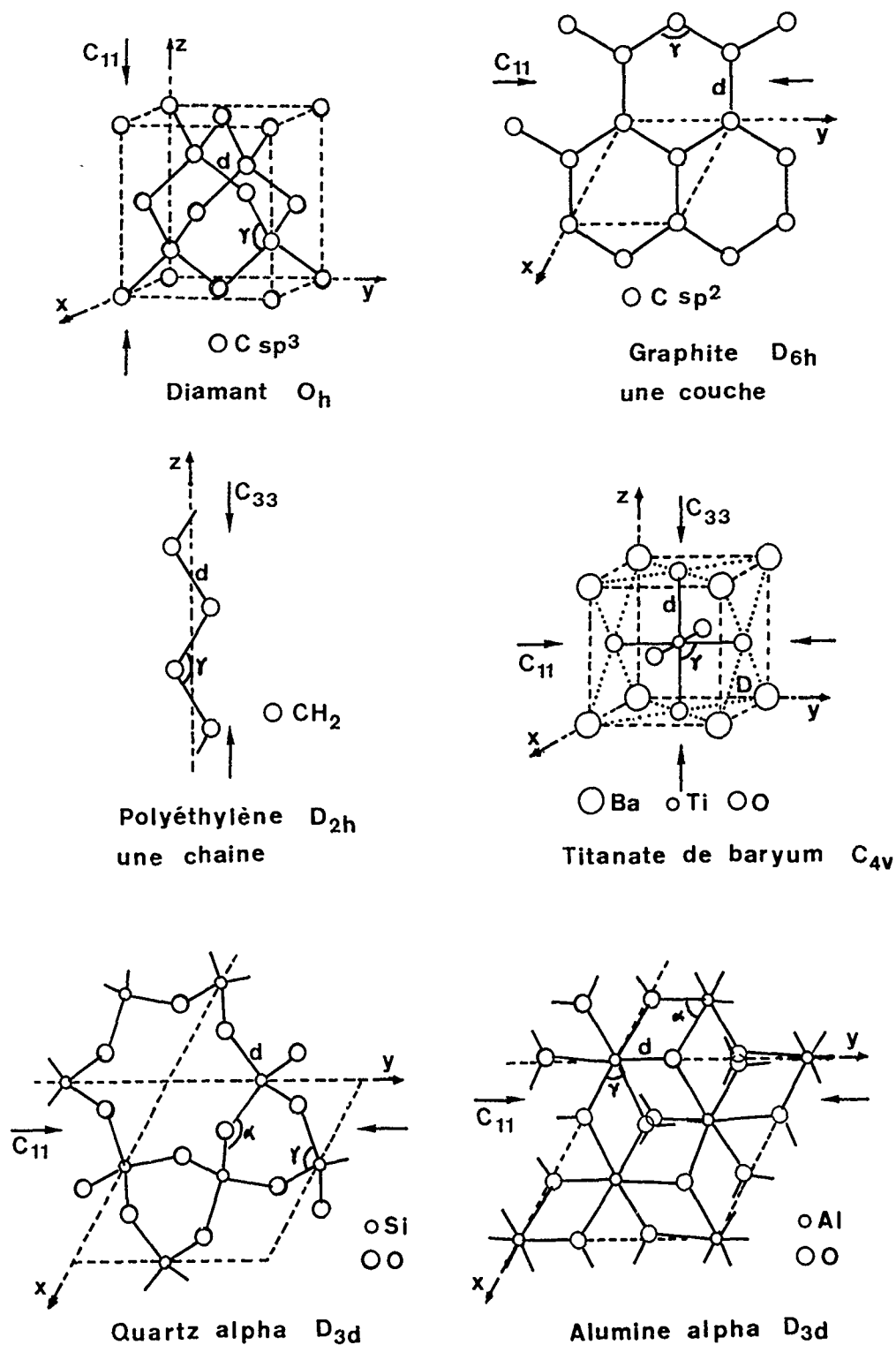


Fig. 1. Structure des composés étudiés avec la définition des principales coordonnées internes introduites.

mentales des fréquences de vibration et des constantes élastiques étaient connues afin (I) d'évaluer la validité de la méthode, notamment pour des composés iono-covalents comme les oxydes et (II) de comprendre le comportement élastique des composés étudiés.

COMPOSES ETUDIES

Nous avons étudié des composés très divers dans leur nature chimique et leur type de symétrie :

- des composés carbonés : C diamant, C graphite (et son composé voisin le nitrure de bore BN), polyéthylène.
- des oxydes : quartz alpha et alumine alpha
- un oxyde mixte, le titanate de baryum, BaTiO_3 .

Pour tous ces composés, nous avons utilisé les données de la littérature concernant l'attribution des fréquences de vibration aux différents modes de symétrie ainsi que les valeurs expérimentales des constantes élastiques.

RESULTATS

Sur la figure 1 sont représentées les structures des composés étudiés avec la définition des principales coordonnées internes introduites. Sur le tableau 1 sont regroupées les valeurs des constantes de force correspondantes ainsi que les valeurs expérimentales des constantes élastiques de compressibilité et de cisaillement et la DEP pour chaque composé.

Les calculs ont montré que :

1°) les champs de force de valence généralisés utilisés sont tout-à-fait valables pour calculer avec une bonne approximation les constantes élastiques et les fréquences de vibration, même dans le cas de composés iono-covalents où existent de nombreuses interactions; les fréquences sont généralement ajustées avec une erreur de l'ordre de 3%, pour les constantes élastiques, l'erreur moyenne est inférieure à 20%, ce qui est satisfaisant si l'on considère que les erreurs les plus grandes interviennent pour les éléments extra-diagonaux du tenseur d'élasticité qui ont en général des valeurs faibles et peu précises.

2°) la contribution relative de chaque type de liaison et de chaque type d'angle aux constantes élastiques étant connue, il est possible de mettre en évidence les paramètres qui influencent le plus le comportement élastique des composés.

Les résultats obtenus pour les composés carbonés montrent que les constantes élastiques sont très sensibles à la dimensionalité du réseau covalent. On observe tout d'abord que les constantes de compressibilité C_{11} du diamant et du graphite et C_{33} du polyéthylène présentent une contribution des elongations des liaisons C-C et des déformations des angles C-C-C, tandis que ces dernières sont prépondérantes dans les constantes de cisaillement.

Si l'on compare les résultats obtenus, on observe que :

- les constantes C_{11} du diamant (tridimensionnel, 3D) et du graphite (2D) sont égales, alors que les constantes de force f_d et f_γ sont environ deux fois plus grandes dans le graphite que dans le diamant.

- la constante C_{11} du diamant (3D) est environ quatre fois plus grande que la constante C_{33} du polyéthylène (1D), bien que f_d et f_γ soient deux fois plus grandes dans le polyéthylène que dans le diamant.

On peut donc dire qu'à constantes de force comparables, les constantes élasti-

TABLEAU 1

Constantes de force principales, constantes de compressibilité et de cisaillement et DEP pour les composés étudiés.

composés	constantes de force définition	valeurs ^a	C_{ij}	constantes élastiques valeurs ^b	DEP ^c
C diamant	d(C-C)	3,80	C_{11}	1,06	$\underline{d} + \gamma$
O_h	γ (C-C-C)	0,51	C_{44}	0,57	$\underline{d} + \underline{\gamma}$
C graphite	d(C-C)	6,88	C_{11}	1,06	$\underline{d} + \gamma$
	γ (C-C-C)	1,08	C_{33}	0,036	\underline{l}
D_{6h}	l(liaison intercouche)	0,014	C_{44}	0,004	\underline{l}
			C_{66}	0,44	γ
Polyéthylène	d(C-C)	4,71	C_{11}	0,01	liaisons Van der Waals
C_{2h}	γ (C-C-C)	1,10	C_{22}	0,009	
			C_{44}, C_{55}, C_{66}	< 0,003	
			C_{33}	0,235	
					$\underline{d} + \gamma$
Quartz α	d(Si-O)	5,77	C_{11}	0,09	$\gamma + \alpha$
	γ (O-Si-O)	0,75	C_{33}	0,107	$\gamma + \alpha$
D_{3d}	α (Si-O-Si)	0,22	C_{44}	0,06	$\underline{d} + \underline{\gamma} + \alpha$
			C_{66}	0,04	$\underline{\gamma} + \alpha$
Alumine α	d(Al-O)	1,15	C_{11}	0,50	$\underline{d} + \gamma + \alpha$
	γ (O-Al-O)	0,48	C_{33}	0,50	$\underline{d} + \underline{\gamma} + \alpha$
D_{3d}	α (Al-O-Al)	0,28	C_{44}	0,15	$\underline{d} + \gamma + \alpha$
			C_{66}	0,17	$\underline{d} + \underline{\gamma} + \alpha$
BaTiO ₃	d(Ti-O)	0,87	C_{11}	0,275	$\underline{d} + D$
	γ (O-Ti-O)	0,42	C_{33}	0,165	$\underline{d} + D$
C_{4v}	D(Ba-O)	0,19	C_{44}	0,054	$\gamma + D$
			C_{66}	0,113	$\gamma + D$

^aélongations de liaisons en mdyn.Å^{-1} ; déformations angulaires en mdyn.Å.rad^{-2} .

^ben mdyn/Å^2

^cdistribution d'énergie potentielle (voir texte)

ques de compressibilité augmentent avec la dimensionalité du réseau covalent.

Les résultats obtenus pour le titanate de baryum montrent l'influence de l'orientation des liaisons par rapport aux axes du cristal. Dans ce composé, la charpente est constituée d'octaèdres TiO_6 associés par les sommets, le réseau Ba-O, beaucoup plus ionique est peu rigide. Les constantes de compressibilité comportent une très forte contribution des élongations des liaisons Ti-O qui sont parallèles aux axes du cristal, tandis que les constantes de cisaillement ont une contribution des déformations angulaires O-Ti-O et des élongations des liaisons Ba-O. On peut donc dans ce cas, relier directement la rigidité des liaisons Ti-O aux constantes de compressibilité.

Enfin, l'étude du quartz α et de l'alumine α montre l'importance de la coordinence des atomes dans le comportement élastique d'un composé. Dans le quartz, le silicium et l'oxygène ont respectivement une coordinence 4 et 2, dans l'alumine, l'aluminium a une coordinence 6 et l'oxygène une coordinence 4. On constate, dans le cas du quartz, que l'élongation des liaisons Si-O ne joue pratiquement aucun rôle dans les constantes de compressibilité comme de cisaillement; ceci explique les faibles valeurs de C_{11} et C_{33} : la structure est considérée comme flexible (ref. 5) malgré la forte rigidité des liaisons Si-O. La faible coordinence des atomes d'oxygène est à l'origine de cette flexibilité : lors d'une compression, les angles Si-O-Si se déforment facilement, entraînant la déformation des angles O-Si-O, mais laissant les liaisons Si-O inchangées. Ces résultats sont en bon accord avec des études structurales réalisées sous pression par Jorgensen (refs 5,6) et Le Page et coll (ref. 7).

Par contre, dans l'alumine α où la coordinence des atomes est plus élevée, le réseau cristallin est difficilement déformable. On observe que l'élongation des liaisons Al-O a une contribution, qui peut être prépondérante, à côté de celle des déformations des angles O-Al-O et Al-O-Al, aux constantes de compressibilité et de cisaillement. Ceci explique que les constantes de compressibilité de l'alumine α sont environ cinq fois plus grandes que celles du quartz α , alors que la constante de force des liaisons Al-O est environ cinq fois plus petite que celle des liaisons Si-O.

CONCLUSION

Cette étude a permis de relier l'élasticité macroscopique et l'élasticité microscopique dans des composés chimiquement très différents. On a pu, pour chacun d'eux, déterminer un champ de force de valence généralisé permettant une bonne approche à la fois des fréquences de vibration et des constantes élastiques. Le développement futur d'un tel travail est la prévision du comportement élastique de matériaux pour lesquels les éléments C_{ij} du tenseur d'élasticité ne sont pas connus, à partir des données structurales et de champs de force déduits des données de spectroscopie vibrationnelle.

REFERENCES

- 1 E.B Wilson, J.C. Decius and P.C. Cross, Molecular vibrations, the theory of infrared and Raman vibrational spectra. Mc Graw Hill Ed. 1955.
- 2 T. Shimanouchi, M. Tsuboi and T. Miyazawa, The J. of Chem. Phys. 35,1597,1961
- 3 Y. Shiro and T. Miyazawa, Bull. of the Chem. Soc. of Japan, 44,2371,1971.
- 4 M. Born and K. Huang, Dynamical Theory of Crystal Lattices. Oxford University Press, London. 1954.
- 5 J.D. Jorgensen, J. Appl. Phys. 49,5473,1978.
- 6 L. Cartz and J.D. Jorgensen, J. Appl. Phys. 52,236,1981.
- 7 Y. Le Page, L.D. Calvert and E.J. Gabe, J. Phys. Chem. Solids, 41,721,1980.

A SIMPLE MODEL OF NUCLEIC BASES ASSOCIATION : THE CYCLIC DIMER OF THE 2-PYRIDONE
VIBRATIONAL STUDY*

J. FAVROT, Th. BOUISSOU, J.F. BRAZIER¹, S. CASTILLO² and A. ZWICK³

¹U.A. au CNRS 454, Laboratoire Synthèse, Structure et Réactivité de Molécules
Phosphorées, Université Paul Sabatier, 118 Route de Narbonne 31062 TOULOUSE
CEDEX - FRANCE

²Laboratoire de Recherche sur l'Energie, Université Paul Sabatier,
118 Route de Narbonne 31062 TOULOUSE CEDEX - FRANCE

³U.A. au CNRS 74, Laboratoire de Physique des Solides, Université Paul Sabatier
118 Route de Narbonne, 31062 TOULOUSE CEDEX - FRANCE

SUMMARY

The distinct vibrational frequencies for the 2-pyridone as monomer and centrosymmetric dimer are studied, with respect to concentration in CHCl₃ (or CDCl₃) and CH₃CN (or CD₃CN) solutions. Distinct frequencies have been observed for about one half of the fundamental modes not only for the ones directly involved in hydrogen bond association (modes of the NH and CO groups) but also for ring modes sensitive to the association state. The present results can serve to distinguish a free from an hydrogen bonded nucleic base.

INTRODUCTION

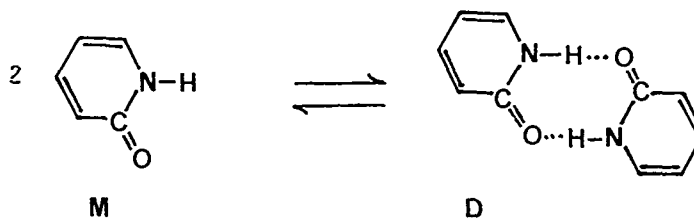
Vibrational spectroscopy, specially ultraviolet resonance Raman spectroscopy, is often used to study nucleic bases association or tautomerism, nucleic acids structural changes or nucleic acids interaction with proteins, metal ions or drugs (ref. 1). For such purposes, it would be of interest to correlate the frequency shifts or the intensity variations of the lines to the association state of the nucleic bases.

The nucleic bases, which possess a pyrimidine or purine skeleton, may give rise to tautomerism involving their hydroxy-(oxo-) or amino-(imino-) group. Such tautomerisms may be correlated with mutagenesis (refs. 2-4) because the association between two bases by hydrogen bond leads to a mispairing when one of the bases is present in a rare tautomeric form. The 2-pyridone molecule is a good one to study such phenomenon because it possesses less heteroatoms than nucleic bases and then less possibilities of association.

For heterocyclic molecules, tautomerism (ref. 5) and infrared spectra (refs. 6-8) have been reviewed. A special review concerns the pyridine derivatives (ref. 9).

*Taken in part from a thesis (Doctorat d'Etat) to be submitted to Paul Sabatier University, Toulouse, France, by S. Castillo. Preliminary results were presented on the occasion of the 100th birthday of Professor G. Mignonac Toulouse 1989 and at the "Journées de Chimie Physique", Paris, May 1989.

The unsubstituted 2-pyridone appears as predominant lactam tautomer in the solid state and in aqueous solution (refs.5,9). In the solid state it occurs in helicoïdal chains (refs.10-11). In various solvents there is an equilibrium between the 2-pyridone monomer (M) and the centrosymmetric cyclic dimer (D) of this molecule :



Many previous papers give data on the 2-pyridone vibrational spectra (refs. 12-22) but only a few have proposed some partial and sometimes conflicting assignments (refs.16,18,21). Isotopic substitution $^{18}\text{O}/^{16}\text{O}$ (refs.19-20), ND/NH and $^{15}\text{N}/^{14}\text{N}$ (ref. 20) has shown that $\nu\text{C} = 0$ and δNH motions are coupled with motions of the ring but experimental data have not been fully interpreted. A same vibrational mode has been assigned at different frequencies for the monomer and the cyclic dimer only for the νNH (refs. 13,15) and $\nu\text{C} = 0$ (ref. 22) modes.

In studies on pyridinium (refs. 23-24) pyrazinium (refs. 23,25) or pyrimidinium (refs. 23,26) salts or on 2-pyrimidone and 2-pyrimidone chlorhydrate (refs. 27-28), some ring frequencies have been found sensitive to the bonding of an hydrogen atom to nitrogen (NH or NH^+). Recent work, with both experimental and theoretical approaches, on heterocyclic molecules such pyridine (refs. 29-30) γ -pyrone (ref. 31), uracile (ref. 32) may also be used to precise the assignment of 2-pyridone vibrational spectra.

Complete assignment of 2-pyridone and N-methyl-2-pyridone vibrational spectra will be published elsewhere (ref. 33). In the present work we emphasize only the distinct frequencies for 2-pyridone monomer or dimer which may serve to distinguish a free from an hydrogen bonded nucleic base, with oxo group.

VIBRATIONS OF THE NH AND CO GROUPS

The NH and CO groups are directly involved in the hydrogen bond and consequently are expected to be more sensitive to the association state.

Modes νNH and νCO (table 1 and figures 1 (νNH) and 2 (νCO)).

It is noticeable that the νNH band of the cyclic dimer (solution of 2-pyridone 0.5 M in CDCl_3) is centered at lower frequency than the νNH band of the solid. Then the N-H...O hydrogen bond is stronger in the centrosymmetric dimer than in the helicoïdal chains of the solid. Similar data have been previously reported in the νNH (refs. 13,15) and the νCO (ref. 22) ranges and interpreted in the same manner.

TABLE 1. Vibration modes of the 2-pyridone sensitive to the association state. Distinct infrared frequencies for monomer and dimer in chloroform and acetonitrile solutions^a.

Solid ^b	Assignment ^c	Solution in CDCl_3 ^d		Solution in CH_3CN ^d		Monomer (M) or Dimer(D)
		0.5 M	10^{-2}M	0.5 M ^e	10^{-2}M	
3071	νNH	<u>3397</u>	<u>3397</u>	<u>3312</u> ^h	<u>3317</u>	M
		<u>2829</u>	<u>2828</u>	<u>2827</u> ^h	<u>2828</u>	D
1652	$\nu\text{C=O}$	<u>1674</u>	<u>1674</u>	<u>1676</u> ^{i,j}	<u>1676</u>	M
		<u>1656</u>	<u>1657</u>	<u>1659</u> ^{i,j}	<u>1659</u>	D
1455	19 a		<u>1474</u> ^{g,j}	<u>1475</u> ^h	f	M
		<u>1472</u>	<u>1471</u> ^j	<u>1472</u> ^h	f	D
1430sh	19 b	<u>1442</u>	<u>1442</u>	<u>1443</u> ^h	f	D
1418			<u>1416</u>	<u>1425</u> ^h	f	M
1363	14	<u>1376</u>	<u>1376</u>	<u>1376</u> ^{h,j}	f	D
		<u>1364</u>	<u>1363</u>	<u>1367</u> ^{h,j}	f	M
1241	δNH	<u>1254</u>	<u>1254</u>	<u>1251</u>		D
			<u>1246</u>	<u>1240</u>	<u>1240</u>	M
1232	δCH	<u>1234</u>	<u>1234</u>	<u>1236</u> ^j	<u>1237</u>	D
		<u>1212</u>	<u>1217</u>	<u>1214</u> ^j	<u>1213</u>	M
1098	δCH	<u>1099</u>	<u>1099</u>	<u>1099</u>		D
			<u>1095</u>	<u>1090</u>	<u>1089</u>	M
981	γNH	<u>995</u>	<u>995</u>	<u>990</u>	<u>991</u>	D, M
845	1	<u>848</u>	<u>846</u>	<u>851</u>	<u>850</u>	D
			<u>830</u>	<u>828</u>	<u>828</u>	M
730	γCH	f	f	<u>730</u>		D
				<u>725</u>	<u>724</u>	M
560	δCO	<u>562</u>	<u>562</u>	<u>560</u>	<u>559</u>	D
			<u>549</u>	<u>549</u>	<u>549</u>	M
475	γCO			<u>512</u>	<u>510</u>	M
		<u>494</u>	<u>495</u>	<u>494</u>		D

a- The concentrations have been selected for the following reasons :
 -for 0.5 M concentration in CHCl_3 the dimer is widely predominant.
 -in 0.2 M solution in CH_3CN or 10^{-2}M in CHCl_3 , for most of the modes the doublet components relative to monomer and dimer have intensities of the same order of magnitude.-for a 10^{-2}M solution in CH_3CN the dimer is widely predominant.
 b- KBr pellet. Many shoulders appear on the solid state spectrum of the 2-pyridone. It is sometimes difficult to link the solid and solutions spectra. The assignment in terms of monomer and dimer works only for spectra of 2-pyridone solutions.
 c- Ring modes with benzene notation (ref. 23). d- For a doublet relative to the same vibration mode of the monomer and of the dimer we have underlined the frequency of the more intense component. e- Except otherwise mentioned. f- masked by solvent band. g- Uncertain, partially masked by solvent band (poor compensation). h - 2-pyridone 0.2 M in CD_3CN . i- 2-pyridone 0.2 M in CH_3CN . j- Nearly equal intensity for monomer and dimer components for this concentration.

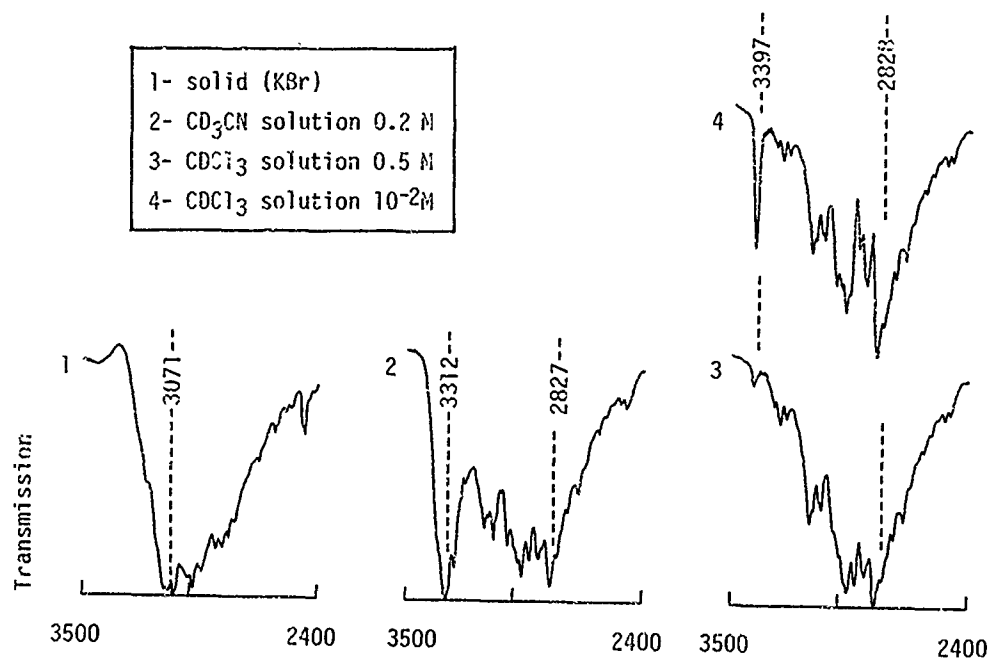


Fig. 1 Infrared spectra of the 2-pyridone in solid state and in solution. ν_{NH} range.

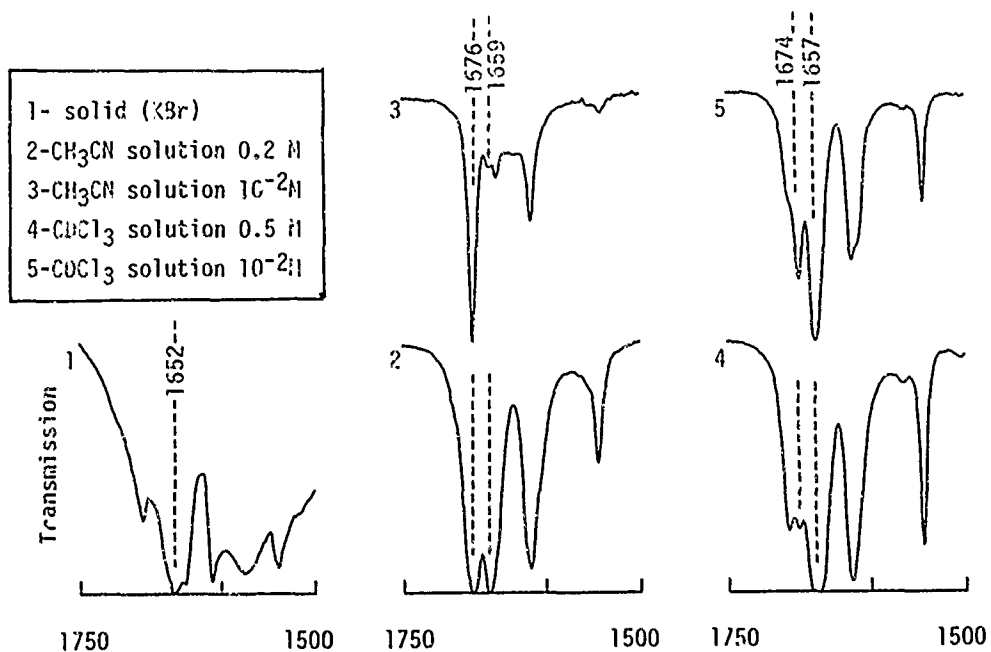


Fig. 2 Infrared spectra of the 2-pyridone in solid state and in solution. $\nu_{\text{C=O}}$ range.

Modes δNH and γNH (table 1 and figure 3 (δNH))

For these two modes, our assignment is coherent with the previous ones for pyridinium salts (refs. 23-24), 2-pyrimidone and its chlorhydrate (refs. 27-28) and with the $^{15}\text{N}/^{14}\text{N}$ isotopic shifts observed for the 2-pyridone (ref. 20). Furthermore this assignment is supported by the fact that the strong and large infrared bands at 1241 and 981 cm^{-1} for 2-pyridone in the solid state are missing in the infrared spectrum of pure N-methyl-2-pyridone (ref. 33). The dilution effect, both in CH_3CN and CDCl_3 , shows clearly that the δNH mode appear near 1250 cm^{-1} (D) and near 1240 cm^{-1} (M). The major polarized Raman lines, for solutions of 2-pyridone 0.2 M in CH_3CN or CD_3CN , lie at 1252 (D) and 1239 cm^{-1} (M). In 1260-1200 cm^{-1} range, one δCH mode is also observed near 1235 cm^{-1} (D) and near 1215 cm^{-1} (M). For the γNH mode it has not been possible to observe at different frequencies the γNH modes of the dimer and of the monomer.

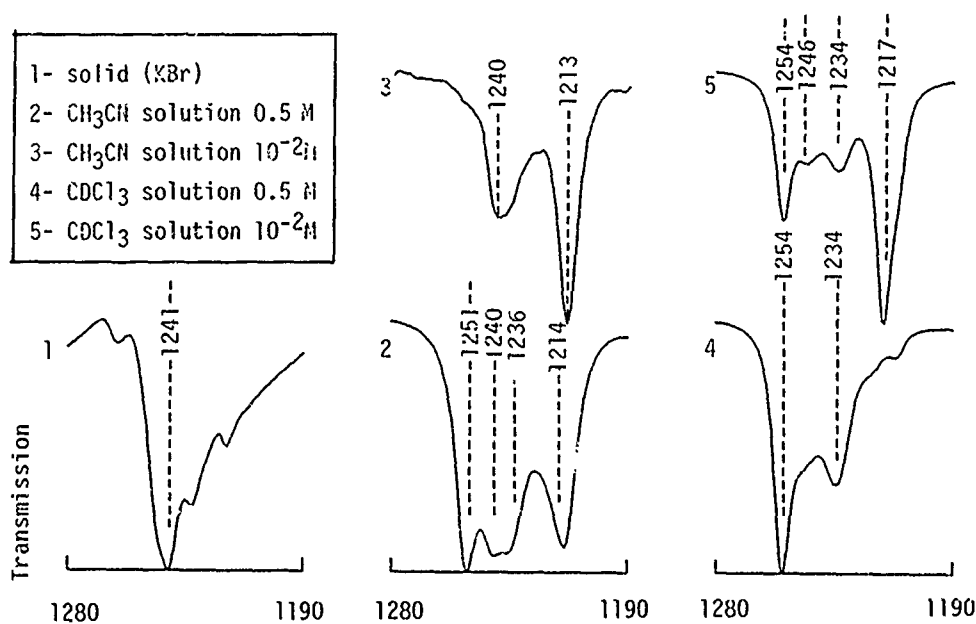


Fig. 3 Infrared spectra of the 2-pyridone in solid state and in solution. 1280-1190 cm^{-1} range.

Modes δCO and γCO (table 1 and figure 4)

Our interpretation, supported by dilution effects in CH_3CN and CHCl_3 , is coherent with the $^{18}\text{O}/^{16}\text{O}$ isotopic shifts observed for the 2-pyridone (ref. 19) and consistent with studies on 2-pyrimidone (refs. 27-28) and 2-pyridinethione (ref. 34).

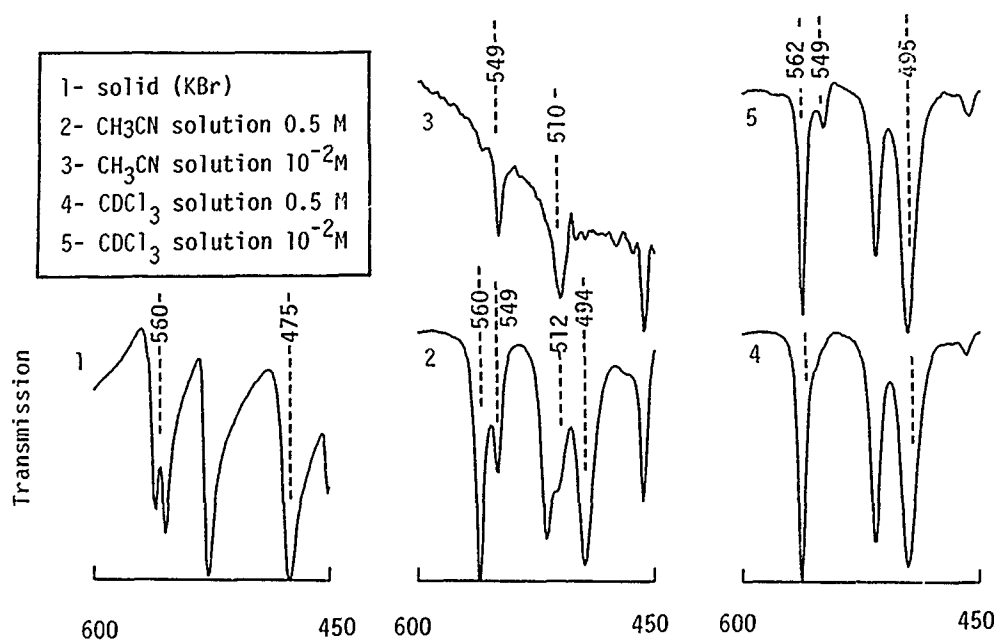


Fig. 4 Infrared spectra of the 2-pyridone in solid state and in solution. δ CO and γ CO ranges.

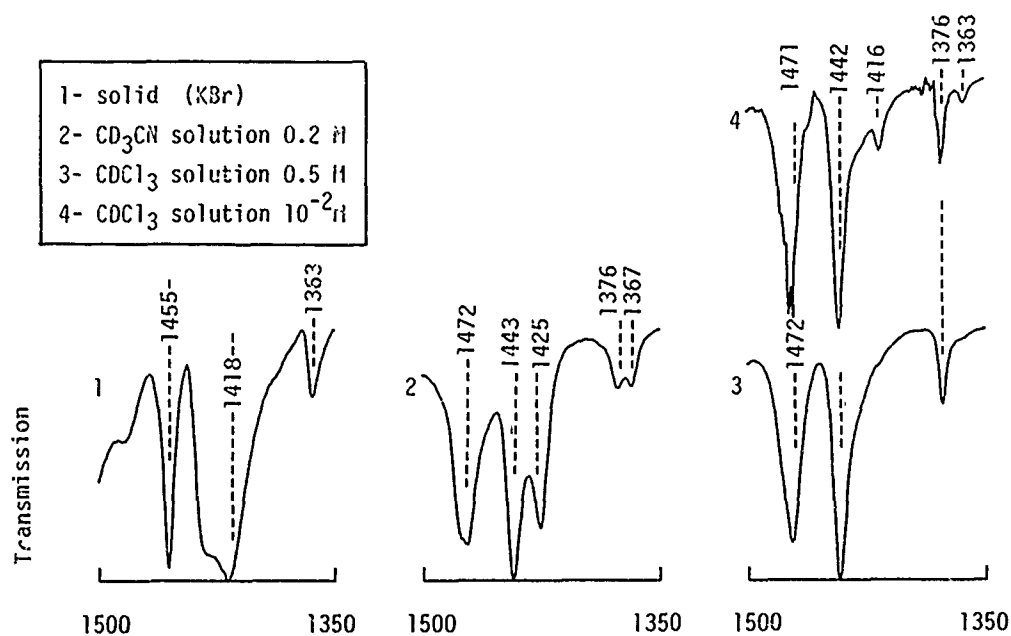


Fig. 5 Infrared spectra of the 2-pyridone in solid state and in solution. 1500-1350 cm⁻¹ range.

RING VIBRATIONS SENSITIVE TO THE ASSOCIATION STATE

With the help of dilution effects in deuterated solvents (CD_3CN or CDCl_3) it has been possible to assign distinct frequencies to monomer and dimer for the 19a, 19b, 14 ring modes in the $1500\text{--}1350\text{ cm}^{-1}$ range (table 1 and figure 5). Such data have not been previously reported because previous studies in solution have been made only in hydrogenated solvents.

Other distinct frequencies for monomer and dimer (table 1) are one δCH near 1100 cm^{-1} , the 1 ring mode at 850 cm^{-1} , and one γCH around 730 cm^{-1} . In solution of 2-pyridone 0.2 M in CH_3CN strong and polarized Raman lines are observed at 843 (D) and 826 cm^{-1} (M) for the 1 ring mode.

REFERENCES

- 1 Proceedings of the eleventh International Conference on Raman Spectroscopy, 5-9 September 1988, London, England Ed. by R.J.H. Clark and D.A. Long, John Wiley and sons, 1988.
- 2 F. Chapeville, H. Clauser et al, *Biochimie*, Hermann, Paris, 1974, pp. 539-540 and 807-808.
- 3 L. Stryer, *Biochemistry*, W.H. Freeman and Cie, San Francisco, 1975, pp. 640-641.
- 4 J.D. Watson, *Biologie moléculaire du gène*, 3ème ed., Interéditions, Paris, 1978, pp. 252-257 (French translation of "Molecular Biology of the Gene", Third Edition, Benjamin, Menlo Park, 1976).
- 5 J. Elguero, Cl. Marzin, A.R. Katritzky, P. Linda, *Advances in Heterocyclic Chemistry*, Ed. by A.R. Katritzky and A.J. Boulton, Supp. 1. The tautomerism of heterocycles. And references therein-Academic Press, New York 1976.
- 6 A.R. Katritzky, *The infrared spectra of heteroatomic compounds*, Quart-Rev. (London), 13 (1959) 353-373.
- 7 A.R. Katritzky and A.P. Ambler, in *Physical Methods in Heterocyclic Chemistry*, A.R. Katritzky, ed. vol II, Academic Press, New York, 1963, ch. 10 Infrared Spectra, pp 161-360.
- 8 A.R. Katritzky and P.J. Taylor, in *Physical Methods in Heterocyclic Chemistry*, A.R. Katritzky, ed. vol IV, Academic Press, New York, 1971, ch 6. Infrared spectroscopy of heterocycles p. 265-434.
- 9 H. Tieckelmann, *The Chemistry of Heterocyclic Compounds*, vol. 14, A. Weissberger and E.C. Taylor eds., Pyridine and its derivatives supplement, R.A. Abramovitch ed., Part Three-John Wiley and Sons, 1974, ch. 12, Pyridinols and Pyridones, pp. 597-1180.
- 10 B.R. Penfold, *The electron distribution in crystalline α -pyridone*, *Acta Cryst.* 6 (1953) 591-600.
- 11 U. Ohms, H. Guth, E. Hellner, H. Dannohl and A. Schweig, 2-pyridone, $\text{C}_5\text{H}_5\text{NO}$, crystal structure refinements at 295 K and 120 K, experimental and theoretical deformation density studies, *Z.Kristallogr.* 169 (1984) 185-200.
- 12 J.A. Gibson, W. Kynaston and A.S. Lindsey, *The infrared spectra of some pyridones and quinolones and their behaviour in the Kolbe-Schmitt reaction*, *J. Chem. Soc.* (1955) 4340-4344.
- 13 Hideyo Shindo, *Studies on the infrared spectra of heterocyclic compounds VI Infrared spectra of substituted α -pyridones and α -quinolones*, *The region from 2000 to 4000 cm^{-1}* , *Chem. Pharm. Bull. Tokyo* 7 (1959) 407-416.
- 14 A. Albert and E. Spinner, *The vibration spectra and structures of the hydroxy-pyridines, and hydroxy-pyrimidines in aqueous solutions*, *J. Chem. Soc.* (1960) 1221-1226.
- 15 L.J. Bellamy and P.E. Rogasch, *Proton transfer in hydrogen bonded systems*, *Proc. Roy. Soc. (London) Ser. A* 257 (1960) 98-108.

- 16 A.R. Katritzky and R. A. Jones, Infrared absorption of heteroatomic and benzenoid six-membered monocyclic nuclei. Part X-Pyridones and pyrid-thiones J. Chem. Soc. (1960) 2947-2953.
- 17 R. Isaac, F.F. Bentley, H. Sternglanz, W.C. Coburn, Jr. C.V. Stephenson and W.S. Wilcox, The far infrared spectra of monosubstituted pyridines, Appl. Spectrosc. 17 (1963) 90-97.
- 18 E. Spinner and J.C.B. White, Spectral and ionisation constant studies of substituted 2-hydroxypyridines (1,2-Dihydro-2-oxypyridines), J. Chem. Soc. (B) (1966) 991-995.
- 19 G.H. Keller, L. Bauer and C.L. Bell, Infrared spectra of 2-pyridone ¹⁶O and 2-pyridone ¹⁸O, Can. J. Chem. 46 (1968) 2475-2479.
- 20 R.A. Coburn and G.O. Dudek, Spectroscopic studies of isotopically substituted 2-pyridones, J. Phys. Chem. 72 (1968) 1177-1181.
- 21 J. Morcillo, M. Gil, D. Escolar, Espectros infrarrojos de oxo e hidroxi-derivados de la piridina, An. Quim. 74 (1978) 1193-1198.
- 22 A. Fujimoto, K. Inuzuka and Ryuichi Shiba, Electronic properties and $\pi-\pi^*$ absorption spectrum of 2-pyridone, Bull. Chem. Soc. Jpn. 54 (1981) 2802-2806.
- 23 R. Foglizzo, Spectres de vibration de quelques sels de pyridinium, pyrazinium et pyrimidinium entre 3300 et 30 cm^{-1} , Thèse de Doctorat ès Sciences Physiques, 1970, Paris.
- 24 R. Foglizzo et A. Novak, Spectres de vibration de quelques halogénures de pyridinium, J. Chim. Phys. 66 (1969) 1539-1550.
- 25 R. Foglizzo and A. Novak, Infrared and Raman spectra of pyrazinium halides, Appl. Spectrosc. 24(6) (1970) 601-605.
- 26 R. Foglizzo et A. Novak, Influence de la protonation et de la complexation sur les spectres de vibration de la pyrimidine, Spectrochim. Acta 26 A (1970) 2281-2292.
- 27 E. Picquenard, Etude du noyau de la pyrimidone-2, mécanisme d'échange hydrogène-deutérium dans la triméthyl-4, 5, 6 pyrimidone-2, spectrométrie de vibration, Thèse de Doctorat ès Sciences Physiques, 1982, Paris.
- 28 E. Picquenard et A. Lautié, Etude par spectrométrie Infrarouge et Raman de la pyrimidone-2 et de son chlorhydrate (chlorure d'oxo-2 pyrimidinium), Spectrochim. Acta 38A(6) (1982) 641-648.
- 29 G. Pongor, P. Pulay, G. Fogarasi and J.E. Boggs, Theoretical prediction of vibrational spectra-1. The in plane force field and vibrational spectra of pyridine, J. Am. Chem. Soc. 106 (1984) 2765-2769.
- 30 G. Pongor, G. Fogarasi, J.E. Boggs and P. Pulay, Theoretical prediction of vibrational spectra : The out-of-plane force field and vibrational spectra of pyridine, J. Mol. Spectrosc. 114 (2) (1985) 445-453.
- 31 P. Csaszar, A. Csaszar, A. Somogyi, Z. Dinya, S. Holly, M. Gal and J.E. Boggs, Vibrational spectra, scaled quantum-mechanical (SQM) force field and assignments for 4H-pyran-4-one, Spectrochim. Acta, 42 A (4) (1986) 473-486.
- 32 L. Harsanyi, P. Csaszar, A. Csaszar, J.E. Boggs, Interpretation of the vibrational spectra of matrix-isolated uracil from scaled ab initio quantum Mechanical Force Fields, Int. J. Quantum chem. 29(4) (1986) 799-815.
- 33 S. Castillo, Th. Bouissou, J.F. Brazier, J. Favrot and A. Zwick, Assignment of the vibrational spectra of 2-pyridone in the solid state and in solution as centrosymmetric dimer. Comparison with N-methyl-2-pyridone, in preparation.
- 34 A. Lautié, J. Hervieu et J. Belloc, Spectres de vibration de la 2-pyridine-thione, de la 3-pyridazinethione et du 2,4-dithiouracile, Spectrochim. Acta, 39A(4) (1983) 367-372.

KINETIC MODELLING OF HETEROGENEOUS-CATALYZED REACTIONS
WITH THE ANACIN SOFTWARE
APPLICATION TO THE HYDRODENITROGENATION OF PHENANTHRIDINE

J. JOFFRE¹, P. GENESTE¹, A. GUIDA¹, G. SZABO² and C. MOREAU¹

¹Laboratoire de Chimie Organique Physique et Cinétique Chimique Appliquées, URA CNRS 418, Ecole Nationale Supérieure de Chimie de Montpellier, 8 rue Ecole Normale, 34075 Montpellier Cedex (France)

²Total-France, BP 27, 76700 Harfleur (France)

SUMMARY

The hydrodenitrogenation of heavy hydrocarbons is represented by a test consisting of the conversion of alkyylanilines in the presence of nitrogenated polyaromatics compounds. These latter compounds, such as phenanthridine, strongly inhibit the conversion of alkyylanilines.

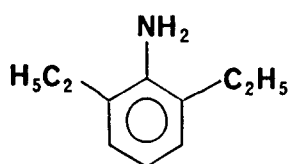
In order to understand these inhibiting effects, a computer tool was required allowing to model the reactivity of the inhibitor. A first order preprogrammed kinetic model was defined taking into account the forward reactions, the reverse reactions and the adsorption constants of the reactants. The constants computed by numerical integration were optimized with the simplex algorithm. In the case of the phenanthridine hydrodenitrogenation, the best fit was obtained with the optimization of all the constants.

The Anacin software thus allows to determine automatically up to twenty constants from one set of experimental data.

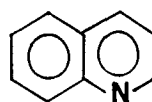
INTRODUCTION

Catalytic hydrodenitrogenation (HDN) is one of the key steps of heavy feedstocks upgrading through their conversion into lighter products. Heavy ends have high nitrogen contents and their conversion yields again distillates too rich in nitrogen. They will be integrated in a normal refining scheme only after a severe HDN treatment (ref. 1). Indeed, combustion of N-compounds produces nitrogen oxides and the presence of nitrogen also leads to the poisoning of metal and acid catalysts used in reforming, cracking and hydrocracking operations. Removal of nitrogen is therefore an absolute necessity for both ecological and economical reasons.

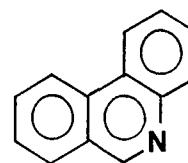
The major problem in the hydrodenitrogenation of distillates resulting from the conversion of heavy feedstocks is the conversion of basic compounds, alkyylanilines in particular, in the presence of other compounds found in the feed. Alkyylanilines appear to be more resistant to hydrodenitrogenation compared to their conversion when used as single components. We have recently developed a simple test to simulate this inhibiting effect on a laboratory scale (ref. 2). This test consists of the conversion of 2,6-diethylaniline in the presence of quinoline or phenanthridine.



2,6-diethylaniline



quinoline



phenanthridine

We have shown, in particular, that the inhibiting effect results from the presence of aromatic, partially or totally saturated polycyclic compounds. This inhibiting effect is less pronounced at low inhibitor concentration. This effect also depends on the progress in the conversion of the inhibitor: the inhibition disappears when the inhibitor is transformed in lighter molecules. This well illustrates the problems of competitive adsorption between substituted alkyylanilines and heavy N-heteroatomics present in the feeds.

In order to gain a better knowledge on these complex inhibiting phenomena and their possible quantification, a computer tool was necessary first to interpret the experimental kinetic results and then to model these inhibiting effects.

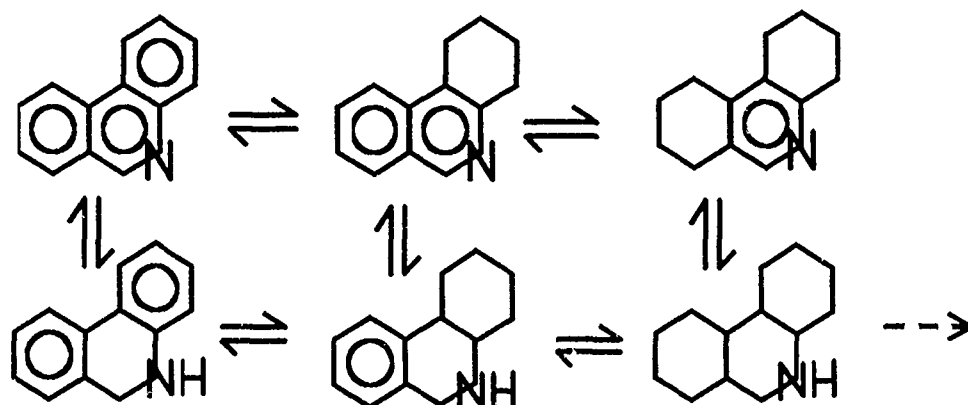
EXPERIMENTAL

Experiments were carried out in a 0.3 litre stirred autoclave operating in a batch mode at 340°C and 70 bars of hydrogen pressure. Analyses were performed on a Girdel 30 gas chromatograph equipped with a flame ionization detector using hydrogen as carrier gas. The columns used were capillary columns, CP Sil 5CB or CP Sil 19CB, 25 m x 0.22 mm. Products were identified by comparison with authentic samples and/or by GC-MS analysis. The catalyst used was Procatalyse HR 346 which has the

following composition : 3% NiO, 14% MoO₃ and 83% Al₂O₃. It was sulfided at atmospheric pressure using a fluidized bed technique with a gas mixture of 15% H₂S and 85% H₂ by volume. The catalyst was heated in flowing H₂/H₂S (gas flow : 120 ml/min) from 20°C to 400°C (8°C/min) and held at 400°C for 4 hours, then cooled and swept with nitrogen.

RESULTS AND DISCUSSION

In Scheme 1 is reported the reaction network we propose for the hydrodenitrogenation of phenanthridine over the NiMo HR 346 catalyst at 340°C and 70 bars H₂. Under these operating conditions the reaction proceeds through successive hydrogenation steps. All these steps are equilibrated as shown by the effect of hydrogen pressure. By increasing the pressure from 70 to 140 bars, the reactions are progressively shifted towards the saturated compounds, in agreement with the law of thermodynamics.

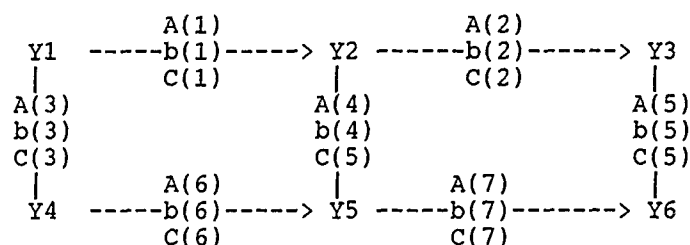


Scheme 1 : Reaction network for the hydrodenitrogenation of phenanthridine at 340°C, 70 bars, over the sulfided NiMo/Al₂O₃ HR 346 catalyst.

This scheme serves us as a frame to model the reactivity of phenanthridine and more generally of complex reaction networks. We have thus developed an interactive software, named AnaCin, to analyse the results of kinetic experiments. This software is easy to use by the average experimental chemist and makes his work easier. AnaCin is written for the Borland's Turbo Basic compiler and uses the 80x87 coprocessor facilities to reduce computation time. This program runs on IBM PC, PS or compatible

computers equipped with a CGA or EGA graphic card, under the MicroSoft Disk Operating System.

The preprogrammed kinetic model used by the software, consisting of six reactants, intermediates or reaction products Y1, Y2 ... Y6, is given in Scheme 2. The A(i) are the rate constants for the forward reactions, C(i) those for the reverse reactions and the b(i) constants are the constants of adsorption for the different compounds (i). In this model the reactions are assumed to be first order in each reactant and all the adsorption sites are supposed to be different.



Scheme 2 : Preprogrammed kinetic model of the AnaCin software

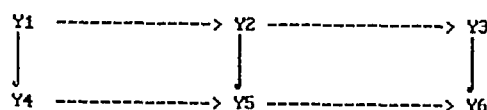
The system of differential equations resulting from the studied model is integrated numerically by the well-known Runge-Kutta method. The constants A(i), C(i) and b(i) are optimized with the simplex method (ref. 3) to fit in with the experimental data.

Application to the hydrodenitrogenation of phenanthridine

The experimental data (Table 1) are entered from the keyboard and they can be corrected, saved or loaded from the disc.

The experimental points are visualized on the screen and the aim is now to obtain the best fit between experimental and computed curves in order to deduce the corresponding rate constants. This can be done manually for simple reaction schemes but an optimization procedure is required for kinetically and chemically more complex reactions.

From the experimental data of Table 1 the optimization on the forward reaction rate constants A(i) gives the results reported in Fig. 1.



N° Exp	Time	Conc1	Conc2	Conc3	Conc4	Conc5	Conc6
1	0	82.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	5	75.0000	1.2000	0.0000	6.2000	0.0000	0.0000
3	10	70.0000	4.3000	0.0000	9.6000	0.0000	0.0000
4	20	60.0000	6.2000	0.0000	10.4000	3.5000	0.0000
5	30	55.0000	9.5000	0.0000	9.2000	5.0000	0.0000
6	40	50.0000	11.5000	0.0000	8.5000	7.0000	0.0000
7	60	40.0000	13.0000	12.1000	7.0000	8.0000	1.0000
8	80	30.0000	15.0000	15.0000	5.7000	8.6000	1.5000
9	100	20.0000	16.6000	21.0000	4.5000	10.0000	2.0000
10	120	15.0000	16.0000	28.0000	2.8000	10.4000	2.2000
11	150	5.0000	15.2000	37.4000	1.8000	9.3000	3.0000
12	200	0.0000	13.0000	45.0000	1.0000	8.0000	3.5000
13	240	0.0000	11.6000	52.3000	0.0000	7.0000	4.2000
14	300	0.0000	0.0000	61.7000	0.0000	5.4000	7.4000

Table 1 : Example of experimental data editing

The optimized residue (weighted sum of squares of errors) is 1228 and it can be seen that some curves do not fit very well, particularly those simulating the reactivity of the intermediates Y2, Y4 and Y5. This was not unexpected because of the reverse reactions present under these experimental conditions (70 bars H₂). A pressure of 140 bars is required to minimize the influence of the reverse reactions.

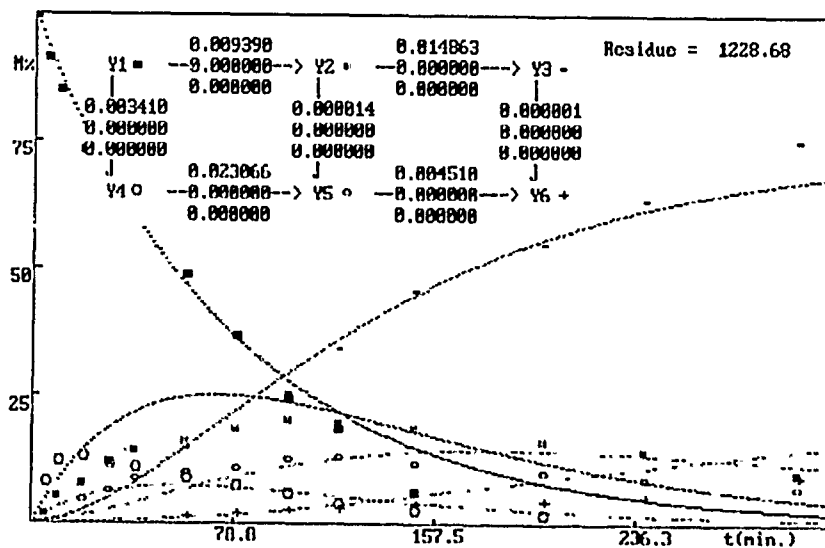


Fig. 1 : Optimization on forward reaction rate constants

A second optimization is then carried out by taking into account the forward and the reverse reactions rate constants. The simulated curves fit fairly well with the experimental data as seen in Fig. 2. The residue is now 388 and these results could be sufficient for the experimenter, considering the experimental error on the determination of the data points. Moreover, the calculated rate constants are consistent with those obtained for related model compounds.

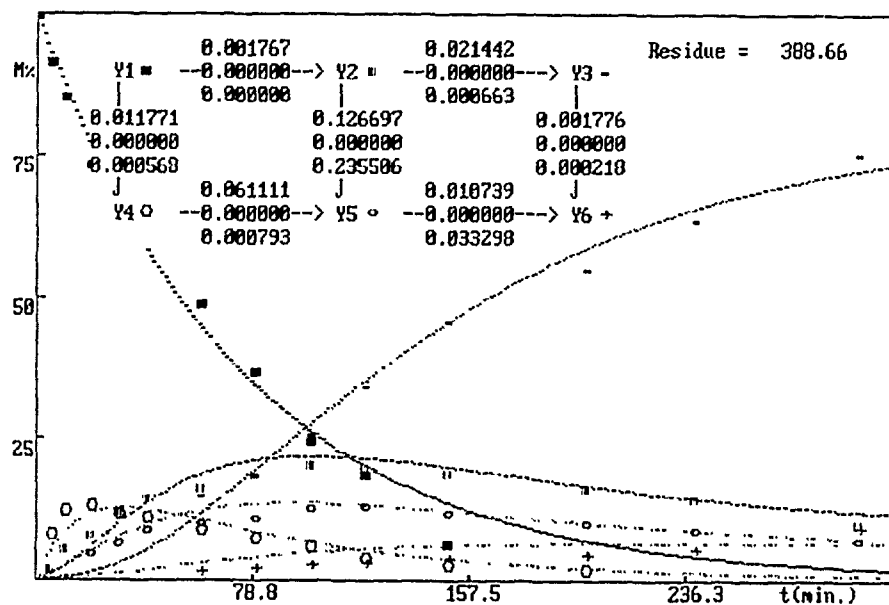


Fig. 2 : Optimization on forward and reverse reaction rate constants

The final optimization (Fig. 3) also taking into account the constants of adsorption of Y1, Y2 ...Y6 yields the smallest residue (250) and the best fit of the simulation curves. These results are still consistent with the reactivity of related model compounds and with their aromatic properties : Y1 and Y4 are the most aromatic and, as a consequence, the most adsorbed, as already observed from other models (ref. 4).

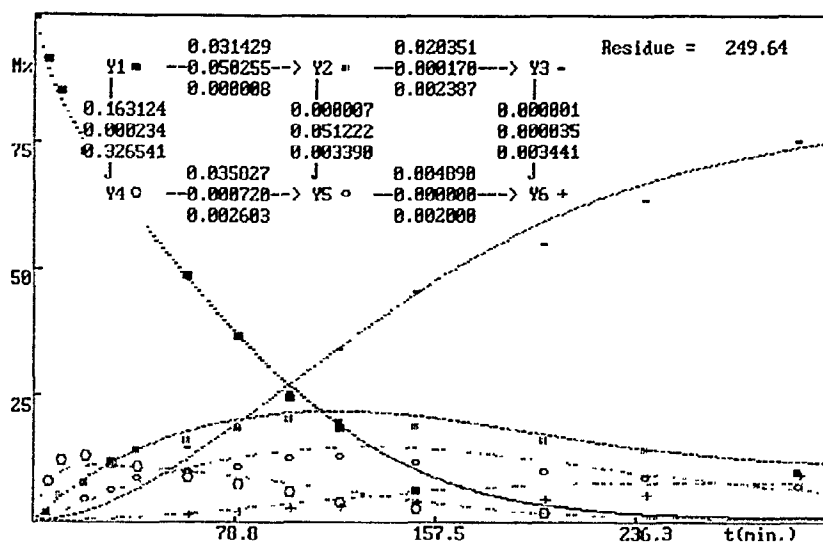


Fig. 3 : Optimization on forward and reverse reaction rate constants and on adsorption constants.

CONCLUSION

The AnaCin software easily allows to interpret the experimental kinetic results obtained for complex reactions. It also confirms the validity of the kinetic model used in our approach. This software works at a reduced running cost and requires only widely used computer equipment. It allows to determine automatically up to twenty constants from one set of experimental data and to visualize the results by comparing experimental and simulated curves.

This first step was an absolute necessity to better understand the inhibition phenomena occurring in hydrotreating reactions. The following step will deal with the modelling of the actual inhibiting effects.

REFERENCES

- 1 H. Toulhoat and R. Kessas, Rev. Fr. I.F.P., 41 (1986) 511.
- 2 C. Moreau, L. Bekakra, R. Durand, N. Zmimita and P. Geneste, Stud. Surf. Sci. Catal., (1989), in press.
- 3 J. A. Nedler and R. Mead, J. Comput., 7 (1965) 308.
- 4 C. Aubert, R. Durand, P. Geneste and C. Moreau, J. Catal., 97 (1986) 169.

DIFFUSION-CONTROLLED REACTIONS. SIMULATION AND NEW THEORETICAL APPROACH

W. DONG, F. BAROS AND J.C. ANDRÉ

Département de Chimie Physique des Réactions, URA328 CNRS,
ENSIC-INPL, 1, rue Grandville, BP451, 54001 Nancy (France)

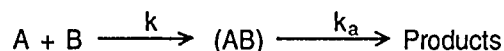
I. INTRODUCTION

Fast chemical reactions in condensed phase are controlled by the transport process of reactants. Traditional approaches for diffusion-controlled reactions are all based on Fick's law to describe the mutual motion of reactants, e.g., the well known Smoluchowski theory (ref. 1,2). Although this kind of approaches have contributed much to our understanding of various diffusion-controlled rate processes, experiments reveal sometimes their inadequacy. For example, under the framework of Smoluchowski theory, the parameters like the encounter diameter and the mutual diffusion coefficient are not consistently determined according to time-resolved or continuous excitation experiments of fluorescence quenching (ref. 3,4). To test stringently the traditional approaches based on diffusion equations and find out the origins of the flaws that they might have, we have recently carried out a molecular dynamics simulation of a model diffusion-controlled reaction in which the solvent is treated on the equal footing of reactants. As can be expected, it is found that the diffusion equation approach is inadequate at short times. Undoubtedly the short time dynamics should not be Markovian. To take into account non-Markovian effect, a new theoretical approach is developed which is based on a generalized diffusion equation. Here we will give a brief review of our recent effort in the study of diffusion-controlled reactions.

II. SIMULATION

A. Model

An irreversible chemical reaction between the reactive solutes A and B in an inert solvent S consists of two steps: the transport of reactants toward each other to form encounter pairs and the intrinsic chemical change which yields final products, i.e.,



where k and k_a are respectively the rate coefficients of the two steps. If the last step is very fast ($k_a \gg k$), the overall rate of the reaction is dominated by the encounter rate

and the reaction is thus called diffusion-controlled, which is the object of the work presented here.

When the intrinsic reaction rate is very large, the chemical transformation of reactants to products can be considered to take place instantaneously once the reactants reach their encounter distance. This is the argument which underlines Smoluchowski's absorbing boundary condition for describing reactivity. In the practice of simulation, this means that each collision between reactive molecules entails certainly an instantaneous reaction.

In our first simulation (ref. 5), reactants and solvent molecules are all modeled by hard spheres of the same size and mass. The only difference of reactants from solvent molecules is their reactivity. In the implementation of a simulation, the distinction of the reactants from the solvent molecules can be made by labelling.

Although the above model for diffusion-controlled reactions is highly idealized, it can be qualified as a simple "civilized" model in which solvent is treated on the equal footing of reactants, in contrast to the continuum solvent model used in the diffusion equation approach.

B. Simulation method

Keeping in mind that we like to study a prototype of fluorescence quenching process, our first simulation is carried out for a special case in which one reactive species A (fluorophores) is highly diluted and the other reactants (quenchers) are in large excess. Under this condition, the reaction of each A is independent of the others. So only one A molecule is needed to generate a reactive event in simulation.

Starting from an equilibrium configuration, reactants A and B are designated randomly. Then the trajectories of all molecules in the simulation system are generated and followed by using the molecular dynamics simulation method (ref. 6). Once the collision between the molecule A and one molecule B occurs, the trajectory generation is stopped and the survival time of the molecule A is recorded. The generation of one reactive event is thus accomplished. To calculate the survival probability of the molecule A, a large number of reactive events is needed, of the order of 10^5 , to obtain a good statistics. A procedure to generate efficiently the reactive events has been developed. The reader is referred to ref. (5) for details.

One main quantity calculated from simulation is the survival probability of the molecule A, which is defined as that the molecule A remains unreacted till to time t after its creation, i.e.,

$$S_A(t) = \lim_{N_A^0 \rightarrow \infty} \frac{N_A(t)}{N_A^0}, \quad (1)$$

where $N_A(t)$ is the number of unreacted molecules A in the ensemble of reactive events at time t and N_A^0 is the total number of reactive events in the ensemble, noting that $N_A^0 = N_A(t=0)$. In practice, $S_A(t)$ is calculated from the directly measured survival times by using

$$S_A(t) = \frac{1}{N_A^0} \sum_{i=1}^{N_A^0} \theta\left(\frac{\tau_i}{t} - 1\right) \quad (2)$$

where τ_i 's are the survival times of the molecule A in reactive events and $\theta(x)$ is Heaviside function.

Once the survival probability is obtained, the time-dependent reaction rate coefficient $k(t)$ can be calculated by using the phenomenological kinetic law

$$\frac{dS_A(t)}{dt} = -k(t) S_A(t) \rho_B \quad (3)$$

where ρ_B is the density of reactants B.

C. Results and discussion

The simulation result of the survival probability is given in Fig. 1. along with the result of Smoluchowski theory which is

$$S_A^S(t) = \exp\left\{-4\pi\sigma D\rho_B\left(1 + \frac{2\sigma}{(\pi Dt)^{1/2}}\right)t\right\} \quad (4)$$

where D is the mutual diffusion coefficient of reactants and σ the encounter diameter. It is seen, from Fig. 1., that the simulation curve of $S_A(t)$ is below that of Smoluchowski theory. This means that Smoluchowski theory underestimates the reaction rate at least in a period of time.

The time-dependent rate coefficient given by Smoluchowski theory is

$$k^S(t) = 4\pi\sigma D\left(1 + \frac{\sigma}{(\pi Dt)^{1/2}}\right) \quad (5)$$

The comparison of this result with simulation is presented in Fig. 2. The most remarkable thing in Fig. 2. is that there is a demarcation of time region. At long times, $t > t_c$, Smoluchowski theory is a good approximation. However, simulation result is quite different from that of Smoluchowski theory at short times, $t < t_c$. The demarcating point time t_c corresponds to about 30 collisions of reactants with solvent molecules before reaction. It is illuminating to note that velocity correlation becomes negligible beyond about 30 collisions in a dense fluid. So the difference observed between

$k^{\text{MD}}(t)$ and $k^{\text{S}}(t)$ is clearly due to dynamic correlations at short times. As consequence the dynamics of reactive encounters is nondiffusional and essentially non-Markovian at short times. In the next section, we will show that the larger reaction rate, compared to Smoluchowski's result, can be accounted for qualitatively by non-Markovian effect.

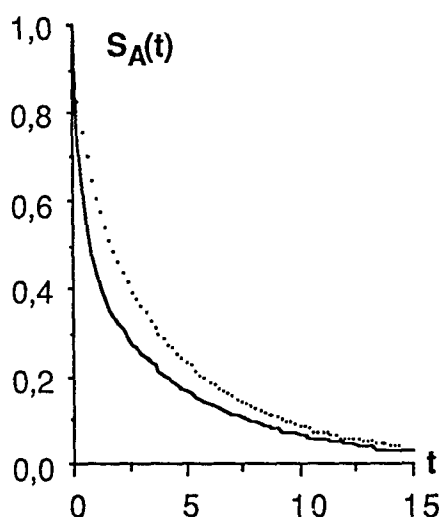


Fig. 1. Survival probability. — MD result ;
 Smoluchowski theory.
 Calculation conditions : total fluid density
 $\rho\sigma^3 = 0.786$ and $\rho_B\sigma^3 = 0.0644$
 Unit for time : $(\sigma^2 m/k_B T)^{1/2}$

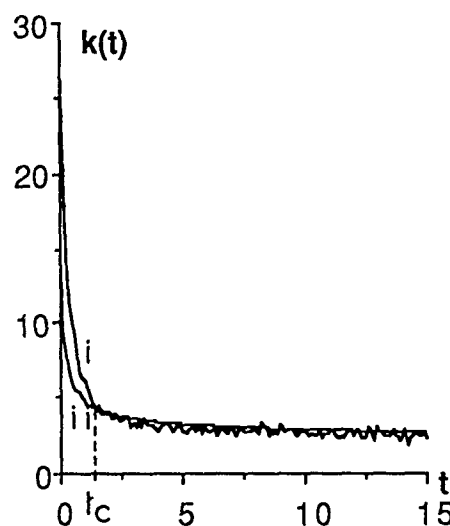


Fig. 2. Time-dependent rate coefficient.
 (i) MD result ; (ii) Smoluchowski theory.
 Same calculation conditions as Fig. 1.
 Unit for rate coefficient : $(\sigma^4 k_B T/m)^{1/2}$

Larger value of reaction rate observed in simulation implies that the exact curve of Stern-Volmer plot is above that of Smoluchowski theory. This has been observed experimentally (ref. 3). Now we can understand at least qualitatively why different results for σ and D are obtained from time-resolved or continuous excitation experiments. The nondiffusional dynamics at short times is wholly contained, in an integrated form, in the results of a continuous excitation experiment. As Smoluchowski theory is inadequate for the short time region, fitting continuous excitation experiment data using this theory will certainly result in some errors for σ and D . Non-Markovian effect seems to be the essential cause.

III. THEORY

From the above discussion, non-Markovian effect is very likely responsible for the nondiffusional behavior of the encounter dynamics at short times. In this section, it is shown how to take it into account and what is its influence on reaction kinetics.

A. Generalized diffusion equation

One convenient starting point of taking non-Markovian effect into account for Brownian motion is the generalized Langevin equation which has the following form for the velocity of a free Brownian particle

$$m \dot{\mathbf{v}}(t) = -m \int_0^t ds \xi(t-s) \mathbf{v}(s) + \mathbf{R}(t) \quad , \quad (6)$$

where m is the mass of the particle, $\mathbf{R}(t)$ the random force and the friction kernel $\xi(t)$ is also called the memory function of the velocity auto-correlation function.

By assuming a Gaussian process for the random force, Adelman has derived, from Eq. (6), a generalized diffusion equation for a free Brownian particle (ref. 7)

$$\frac{\partial \rho(\mathbf{r}, t)}{\partial t} = D(t) \nabla^2 \rho(\mathbf{r}, t) \quad , \quad (7)$$

where $D(t)$ is a time-dependent diffusion coefficient which is related to the memory function by

$$D(t) = \frac{k_B T}{m} L^{-1} [z \hat{\xi}(z)]^{-1} \quad , \quad (8)$$

where k_B is Boltzmann constant, T the absolute temperature and $\hat{\xi}(z)$ is the Laplace transform of the memory function

$$\hat{\xi}(z) = \int_0^\infty dt e^{-zt} \xi(t)$$

and L^{-1} denotes the inverse Laplace transform.

Before using the generalized diffusion equation to treat any concrete problem of diffusion-controlled reactions, the memory function has to be specified so that the time-dependent diffusion coefficient can be then determined. Although there exist now theoretical approaches for calculating memory functions from microscopic theories (ref. 8,9), the relaxation-time approximation is chosen to keep the subsequent development as simple as possible so that a clear insight can be gained into the non-Markovian effect on diffusion-controlled reactions. The adopted memory function takes the following form :

$$\xi(t) = \alpha \delta(t) + \phi_0 e^{-t/\tau} \quad , \quad (9)$$

where α , ϕ_0 and τ are all positive quantities (ref. 10). If $\phi_0 = 0$, the Markovian case is recovered. Moreover it is not difficult to show that the Dirac function part is necessarily present for systems containing hard cores. The approximation of the memory function given in Eq. (9) has also been recently used to study the non-Markovian effect on isomerization reactions (ref. 11).

From Eqs. (8) and (9), straightforward calculations yield

$$D(t) = D(1 + \Delta e^{-t/v}) \quad , \quad (10)$$

where

$$\Delta = \frac{\tau \phi_0}{\alpha} \quad , \quad (11)$$

$$v = \frac{\alpha \tau}{\tau \phi_0 + \alpha} = \frac{\tau}{1 + \Delta} \quad (12)$$

and D is the diffusion constant which is related to α , ϕ_0 and τ through the fluctuation-dissipation theorem

$$D = \frac{k_B T}{m(\alpha + \tau \phi_0)} \quad . \quad (13)$$

The time-dependent diffusion coefficient given in Eq. (10) reduces to its diffusion constant in the limit of long times. At short times the time-dependent part gives a positive contribution. So the transport is faster at short times than that given by the classic Fick's law. It is this enhanced transport which is the origin of the larger encounter rate observed in MD simulation. In the followings, the influence of non-Markovian effect on reaction kinetics is to be examined

B. Generalized Smoluchowski theory

In order to take into account reactivity, we adopt the same scheme as Smoluchowski, i.e., using an absorbing boundary condition at encounter distance. But now the generalized diffusion equation, Eq. (7), will be solved instead of the usual diffusion equation with the following boundary and initial conditions

$$p(r = \sigma, t) = 0 \quad , \quad (14i)$$

$$p(r \rightarrow \infty, t) = 1 \quad (14ii)$$

and

$$p(r, t=0) = \begin{cases} 0 & r < \sigma \\ 1 & r > \sigma \end{cases} \quad . \quad (15)$$

Although Eq. (7) has a nonconstant diffusion coefficient, we have been still able to find the analytic solution under the above conditions.

$$\rho(r, t) = 1 - \frac{\sigma}{r} \operatorname{erfc} \left[\frac{r - \sigma}{\left(\int_0^t ds D(s) \right)^{1/2}} \right], \quad (16)$$

where $\operatorname{erfc}(x)$ is the error function. The method of solution is given in ref. (10).

The time-dependent reaction rate coefficient is then given by

$$\begin{aligned} k(t) &= 4 \pi \sigma^2 D(t) \left. \frac{\partial \rho(r, t)}{\partial r} \right|_{r=\sigma} \\ &= 4 \pi \sigma D(t) \left[1 + \frac{\sigma}{\left(\pi \int_0^t ds D(s) \right)^{1/2}} \right]. \end{aligned} \quad (17)$$

Integrating Eq. (3) and using the above result of the rate coefficient, we find the following result for the survival probability

$$S_A^{\text{GS}}(t) = \exp \left\{ -4 \pi \sigma \rho_B \left[\int_0^t ds D(s) + \frac{2\sigma}{(\pi)^{1/2}} \left(\int_0^t ds D(s) \right)^{1/2} \right] \right\}, \quad (18)$$

where the superscript GS refers to the generalized Smoluchowski theory.

It is worthwhile to note that the above result can be recasted in a very interesting form:

$$S_A^{\text{GS}}(t) = S_A^{\text{S}} \left(\int_0^t ds D(s) \right), \quad (19)$$

where $S_A^{\text{S}}(t)$ is Smoluchowski's result for the survival probability, given in Eq.(4). When its argument is replaced by $\int_0^t ds D(s)/D$, we obtain the result of the generalized Smoluchowski theory which includes the non-Markovian effect.

Now it can be easily shown that the deviation of $S_A^{\text{S}}(t)$ from $S_A^{\text{MD}}(t)$ can be essentially accounted for by non-Markovian effect. Using the relaxation-time approximation for the memory function, we obtain

$$S_A^{\text{GS}}(t) = S_A^{\text{S}}(t + \nu \Delta (1 - e^{-t/\nu})) \quad (20)$$

As can be expected, $S_A^{\text{GS}}(t) \rightarrow S_A^{\text{S}}(t)$ when $t \gg \nu$. But at short times, it can be easily seen that $S_A^{\text{GS}}(t) < S_A^{\text{S}}(t)$. These are qualitatively in agreement with the simulation result (see Fig. 1.). Thus we ascribe the enhanced reaction rate at short times found in MD simulation to non-Markovian effect.

IV. CONCLUSIONS AND PROSPECTS

The results of our recent effort in the study of diffusion-controlled reactions are summarized. A full molecular dynamics simulation has been carried out, in which the solvent is treated on the equal footing of reactants rather than described by a continuum as in traditional approaches, e.g., Smoluchowski theory. It is found that the reactive encounter dynamics is not diffusional at short times, i.e., cannot be adequately described by the classic diffusion equation approach. The demarcation of the time scales is essentially characterized by the duration of dynamic correlations. The nondiffusional dynamics gives an enhanced effect on the reaction rate. The reaction rate found in MD simulation can be about two times that given by Smoluchowski theory at short times. Also this enhanced nondiffusional dynamics plays a key role in understanding the experimental fact that the Stern-Volmer plot measured by a continuous excitation experiment is always above that obtained by integrating the time-dependent fluorescence decay fitted by using Smoluchowski theory.

Dynamic correlations mean that the Markovian description for dynamics is no more adequate. So it is natural that the deviation of Smoluchowski theory from the result of MD simulation at short times is ascribed to non-Markovian effect. To take into account the non-Markovian effect, a new theoretical approach based on a generalized diffusion equation is developed. The non-Markovian effect enters into the theory through a time-dependent diffusion coefficient which is related to the memory function. By using a relaxation-time approximation for the memory function, it can be shown that the motion of the mutual approach of reactive particles is faster than that given by the classic Fick's law. It is this enhanced transport due to dynamic correlations which accounts for satisfactorily the enhanced reaction rate observed in MD simulation. When Smoluchowski's absorbing boundary condition is used for reactivity, the generalized diffusion equation can be still solved analytically. The results turn out to have as simple mathematical expressions as the classic Smoluchowski theory (see e.g. Eqs.(16)-(19)). So the new theoretical approach based on the generalized diffusion equation is strongly recommended for future works on diffusion-controlled reactions, especially for interpreting experimental data. It can be hopefully expected that more consistent parameters can be obtained whether time-resolved or continuous excitation experiment data are used. An analysis of these experimental data by the generalized Smoluchowski theory is planned by our group.

In the last years, the influence of non-Markovian effect on reaction dynamics in condensed phase has attracted much interest (ref. 12). However, theoretical effort has been exclusively focused on the barrier crossing dynamics of activation-controlled reactions. The results presented in this paper shows that the non-Markovian effect is also important for the motion of mutual approach of reactants in diffusion-controlled reactions. The results presented in this paper apply to very simple diffusion-controlled reactions. For many other diffusion-controlled reactions, there appear some

complications, e.g., for reactions between charged species, Coulombic interactions have to be considered and intracellular reactions occur in a limited space. We believe that the non-Markovian effect plays certainly also some role in these diffusion-controlled reactions. The theoretical approach presented in Sec. III. can be extended in a straightforward way to these problems and the results will have mathematical expressions not more complicated than their counterparts in the case without non-Markovian effect.

When a reaction is only partially controlled by diffusion, the reactivity is traditionally treated by the radiation boundary condition (ref. 13). We have made some preliminary attempts to solve the generalized diffusion equation with this boundary condition. But no analytic solution has yet been found. Nevertheless this problem deserves well further effort in expecting to obtain at least approximate analytic solution or upper and lower bounds of the reaction rate by using variational principle.

REFERENCES

1. M.V. Smoluchowski, Z. Phys. Chem., 92, 129 (1917).
2. S.A. Rice, *Diffusion-Limited Reactions*, in Comprehensive Chemical Kinetics Vol. 25, ed. C.H. Bamford, C.F.H. Tipper and R.G. Compton (Elsevier, 1985).
3. T.L. Nemzek and W.R. Ware, J. Chem. Phys., 62, 477 (1975).
4. J.C. André, M. Niclaude and W.R. Ware, Chem. Phys., 28, 371 (1978).
5. W. Dong, F. Baros and J.C. André, J. Chem. Phys. (in press).
6. B.J. Alder and T.E. Wainwright, J. Chem. Phys., 31, 459 (1959).
ibid., 33, 1439 (1960).
7. S.A. Adelman, J. Chem. Phys., 64, 124 (1976).
8. J.P. Hansen and I.R. McDonald, *Theory of Simple Liquids* (2nd edition, Academic Press, 1986).
9. G.F. Mazenko and S. Yip, in *Statistical Mechanics*, Part B, Ed. B.J. Berne (Plenum Press, 1977).
10. W. Dong, F. Baros and J.C. André, Ber. Bunsenges. Phys. Chem. (to appear).
11. S.B. Zhu, J. Lee, G.W. Robinson and S.H. Lin, J. Chem. Phys., 90, 6335 ; 6340 (1989).
12. A. Nitzan, Adv. Chem. Phys., 70, 489 (1988).
13. F.C. Collins and G.E. Kimball, J. Colloid Sci., 4, 425 (1949).

MOLECULAR DYNAMICS : APPLICATIONS TO PROTEINS

M. KARPLUS

Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138,
U.S.A.

INTRODUCTION

Molecular dynamics of macromolecules of biological interest began in 1977 with the publication of a paper on the simulation of a small protein, the bovine pancreatic trypsin inhibitor (McCammon et al., 1977). Although the trypsin inhibitor is rather uninteresting from a dynamical viewpoint--its function is to bind to trypsin--experimental and theoretical studies of this model system--the "hydrogen atom" of protein dynamics--served to initiate explorations in this field.

The most important consequence of the first simulations of biomolecules was that they introduced a conceptual change (Karplus and McCammon, 1981 ; Brooks et al., 1987). Although to chemists and physicists it is self-evident that polymers like proteins and nucleic acids undergo significant fluctuations at room temperature, the classic view of such molecules in their native state had been static in character. This followed from the dominant role of high-resolution x-ray crystallography in providing structural information for these complex systems. The remarkable detail evident in crystal structures led to an image of biomolecules with every atom fixed in place. D.C. Phillips, who determined the first enzyme crystal structure, wrote "the period 1965-75 may be described as the decade of the rigid macromolecule. Brass models of DNA and a variety of proteins dominated the scene and much of the thinking" (Phillips, 1981). Molecular dynamics simulations have been instrumental in changing the static view of the structure of biomolecules to a dynamic picture. It is now recognized that the atoms of which biopolymers are composed are in constant motion at ordinary temperatures. The x-ray structure of a protein provides approximate average atomic positions, but the atoms exhibit fluid-like motions of sizable amplitudes about these averages. Crystallographers have acceded to this viewpoint and have come so far as to sometimes emphasize the parts of a molecule they do not see in a crystal structure as evidence of motion or disorder (Marquart et al., 1980). Thus, the knowledge of protein dynamics subsumes the static picture in

that use of the average positions still allows discussion of many aspects of biomolecule function in the language of structural chemistry. The recognition of the importance of fluctuations opens the way for more sophisticated and accurate interpretations.

Simulation studies on biomolecules have the possibility of providing the ultimate detail concerning motional phenomena (Brooks et al., 1987). The primary limitation of simulation methods is that they are approximate. It is here that experiment plays an essential role in validating the simulations ; that is, comparisons with experimental data can serve to test the accuracy of the calculations and to provide criteria for improving the methodology. When experimental comparisons indicate that the simulations are meaningful, their capacity for providing detailed results often makes it possible to examine specific aspects of the atomic motions far more easily than by making measurements.

In what follows, a brief introduction to molecular dynamics will be given, followed by applications that illustrate its utility for increasing our understanding of proteins, including enzymes, and for interpreting experiments in a more effective way.

METHODOLOGY

To study theoretically the dynamics of a macromolecular system, it is essential to have a knowledge of the potential energy surface, which gives the energy of the system as a function of the atomic coordinates. The potential energy can be used directly to determine the relative energies of the different possible structures of the system ; the relative populations of such structures under conditions of thermal equilibrium are given in terms of the potential energy by the Boltzmann distribution law (McQuarrie, 1976). The mechanical forces acting on the atoms of the systems are simply related to the first derivatives of the potential with respect to the atom positions. These forces can be used to calculate dynamical properties of the system, e.g., by solving Newton's equations of motion to determine how the atomic positions change with time (McQuarrie, 1976 ; Hansen and McDonald, 1976). From the second derivatives of the potential surface, the force constants for small displacements can be evaluated and these can be used to find the normal modes (Levy and Karplus, 1979) ; this serves as the basis for an alternative approach to the dynamics in the harmonic limit (Levy and Karplus, 1979 ; Brooks and Karplus, 1983).

Although quantum mechanical calculations can provide potential surfaces for small molecules, empirical energy functions of the molecular mechanics type are the only possible source of such information for proteins and the surrounding solvent. Since most of the motions that occur at ordinary temperatures leave the bond lengths and bond angles of the polypeptide chains near their equilibrium values, which appear not to vary significantly throughout the protein (e.g., the

standard dimensions of the peptide group first proposed by Pauling [Pauling et al., 1951]), the energy function representation of the bonding can have an accuracy of the order of that achieved in the vibrational analysis of small molecules. Where globular proteins differ from small molecules is that the contacts among nonbonded atoms play an essential role in the potential energy of the folded or native structure. From the success of the pioneering conformational studies of Ramachandran and co-workers (Ramachandran et al., 1963) that made use of hard-sphere nonbonded radii, it is likely that relatively simple functions (Lennard-Jones nonbonded potentials supplemented by electrostatic interactions) can adequately describe the interactions involved.

The energy functions used for proteins are generally composed of terms representing bonds, bond angles, torsional angles, van der Waals interactions and electrostatic interactions. The expression used in the program CHARMM (Brooks et al., 1983) has the form :

$$E(\bar{R}) = 1/2 \sum_{\text{bonds}} k_b (b - b_0)^2 + 1/2 \sum_{\text{bond angles}} K_\Theta (\Theta - \Theta_0)^2 + \quad (1)$$

$$1/2 \sum_{\text{torsional angles}} K_\phi [1 + \cos(n\phi - \delta)] + \sum_{\text{nb pairs } r < 8\text{\AA}} \left(\frac{A}{r^{12}} - \frac{C}{r^6} + \frac{q_1 q_2}{Dr} \right)$$

The energy is a function of the Cartesian coordinate set, \bar{R} , specifying the positions of all the atoms involved, but the calculation is carried out by first evaluating the internal coordinates for bonds (b), bond angles (Θ), dihedral angles (ϕ), and interparticle distances (r) for any given geometry \bar{R} , and using them to evaluate the contributions to Eq. (1), which depend on the bonding energy parameters K_b , K_Θ , K_ϕ , Lennard-Jones parameters A and C , atomic charges q_i , dielectric constant D , and geometrical reference values b_0 , Θ_0 , n , and δ . For most simulations use has been made of a representation that replaces aliphatic group (CH_3 , CH_2 , CH) by single extended atoms. Although the earliest studies employed the extended atom representation for all hydrogens, present calculations treat hydrogen-bonding hydrogens explicitly. In the most detailed simulations every protein atom (including aliphatic hydrogens) and explicit solvent molecules (e.g., a three-site or five-site model for each water molecule) is included (Brooks et al., 1983).

Given a potential energy function, one may take any of a variety of approaches to study protein dynamics. The most detailed information is provided by molecular dynamics simulations, in which one uses a computer to solve the Newtonian equations of motion for the atoms of the protein and any surrounding

solvent (McCammon et al., 1977 ; McCammon et al., 1979 ; van Gunsteren and Karplus, 1982). With currently available computers, it is possible to simulate the dynamics of small proteins for periods of up to a nanosecond. Such periods are long enough to characterize completely the librations of small groups in the protein and to determine the dominant contributions to the atomic fluctuations. To study slower and more complex processes in proteins, it is generally necessary to use other than the straightforward molecular dynamics simulation method. A variety of dynamical approaches, such as stochastic dynamics (Chandrasekhar, 1943), harmonic dynamics (Levy and Karplus, 1979 ; Brooks and Karplus, 1983), and activated dynamics (Northrup et al., 1982), can be introduced to study particular problems (Brooks et al., 1987).

Since molecular dynamics simulations have been used most widely for studying protein motions, we briefly describe the methodology. To begin a dynamical simulation, one must have an initial set of atomic coordinates and velocities. These are usually obtained from the x-ray coordinates of the protein by a preliminary calculation that serves to equilibrate the system (Brooks et al., 1983). The x-ray structure is first refined using an energy minimization algorithm to relieve local stresses due to non-bonded atomic overlaps, bond length distortions, etc. The protein atoms are then assigned velocities at random from a Maxwellian distribution corresponding to a low temperature, and a dynamical simulation is performed for a period of a few psec. The equilibration is continued by alternating new velocity assignments (chosen from Maxwellian distributions corresponding to successively increased temperatures) with intervals of dynamical relaxation. The temperature, T , for this microcanonical ensemble is measured in terms of the mean kinetic energy for the system composed of N atoms as

$$1/2 \sum_{i=1}^N m_i \langle v_i^2 \rangle = \frac{3}{2} N k_B T \quad (2)$$

where m_i and $\langle v_i^2 \rangle$ are the mass and average velocity squared of the i th atom, and k_B is the Boltzmann constant. Any residual overall translational and rotational motion for an isolated protein can be removed to simplify analysis of the subsequent conformational fluctuations ; in a solution simulation, the protein can diffuse through the solvent. The equilibration period is considered finished when no systematic changes in the temperature are evident over a time of about 10 psec (slow fluctuations could be confused with continued relaxation over shorter intervals). It is necessary also to check that the atomic momenta obey a Maxwellian distribution and that different regions of the protein have the same average temperature. The actual dynamical simulation, which provides coordinates and velocities for all the atoms as a function of time, is then

performed by continuing to integrate the equations of motion for the desired time period. The available simulations for proteins range from 25 to 300 psec. Several different algorithms for integrating the equations of motion in Cartesian coordinates are being used in protein molecular dynamics calculations. Most common are the Gear predictor-corrector algorithm, familiar from small molecule trajectory calculations (McCammon et al., 1979) and the Verlet algorithm (Verlet, 1967), widely used in statistical mechanical simulations (van Gunsteren and Berendsen, 1977).

INTERNAL MOTIONS AND THE UNDERLYING POTENTIAL SURFACE

For native proteins with a well-defined average structure, two extreme models for the internal motions have been considered. In one, the fluctuations are assumed to occur within a single multidimensional well that is harmonic or quasiharmonic as a limiting case (Karplus and Kushick, 1981 ; Brooks and Karplus, 1983 ; Levitt et al., 1985). The other model assumes that there exist multiple minima or substates ; the internal motions correspond to a superposition of oscillations within the wells and transitions among them (Austin et al., 1975 ; Frauenfelder et al., 1979 ; Levy et al., 1982 ; Swaminathan et al., 1982 ; Brooks and Karplus, 1983 ; Debrunner and Frauenfelder, 1982). Experimental have been interpreted with both models, but it has proved difficult to distinguish between them (Agmon and Hopfield, 1983 ; Ansari et al., 1985).

To characterize the protein potential surface structurally and energetically (Elber and Karplus, 1987), we use a 300 ps molecular dynamics simulation of the protein myoglobin at 300°K ; details of the simulation method have been presented (Levy et al., 1985). Myoglobin was chosen for study because it has been examined experimentally by a variety of methods and the two motional models have been applied to it (Austin et al., 1975 ; Frauenfelder et al., 1979 ; Levy et al., 1982 ; Agmon and Hopfield, 1983 ; Bialek and Goldstein, 1985). It is ideally suited for the present analysis, because its well defined secondary structure (a series of α -helices connected by loops) facilitates a detailed characterization of the dynamics.

The topography of the potential surface underlying the dynamics can be explored by finding the local energy minima associated with coordinate sets sequential in time (Stillinger and Weber, 1982 ; Stillinger and Weber, 1984). Thirty-one coordinate sets (one every 10 psec) were selected and their energy was minimized with a modified Newton-Raphson algorithm suitable for large molecules (Brooks et al., 1983). Since the coordinate sets all corresponded to different minima, structures separated by shorter time periods were examined to determine how long the trajectory remains in a given minimum. Seven additional coordinate sets (one every 0.05 psec) were chosen and their behavior on

minimization was examined ; if two coordinate sets converged, they corresponded to the same minimum ; if they diverged they corresponded to different minima. The measure for the distance between two structures is their rms coordinate difference after superposition.

Analysis of the short time dynamics demonstrates that convergence occurs for intervals up to 0.15 ± 0.05 ps. Thus, the 300 ps simulation samples on the order of 2000 different minima ; this is a sizable number but it may nevertheless be small relative to the total (finite) number of minima available to such a complex system in the neighborhood of the native average structure (that is, conformations that are native-like and significantly populated at room temperature). The rms differences among the minimized structures reach a maximum value of approximately 2 Å at about 100 psec. Thus, the difference vector $(\underline{R}_K - \underline{R}_{K'})$, where \underline{R}_K represents the coordinates of all the atoms in a native-like conformation K, is restricted to a volume bounded by a radius of 2 Å.

Comparison of the energies of the minimized structures shows that they vary over about 20°K (40 cal/mole) per degree of freedom. Since this difference in energy between the "inherent" structures (Stillinger and Weber, 1982 ; Stillinger and Weber, 1984) is small, they are significantly populated at room temperature. Further, the large number of such structures sampled by the room temperature simulation suggests that the effective barriers separating them are low and that the protein is undergoing frequent transitions from one structure to another. The fluctuations within a well can be described by a harmonic or quasiharmonic model while the transitions among the wells cannot. Estimates based on the time development of the rms atomic fluctuations for mainchain atoms at room temperature (Swaminathan et al., 1982) indicate that 20 to 30 percent of the rms fluctuations are contributed by oscillations within a well and 70 to 80 percent arise from transitions among wells ; for sidechains the contribution from transitions among the multiple wells is expected to be larger. Since energy differences among some of the wells are small, molecules may be trapped in metastable states at low temperatures, in analogy to third law violations in crystals (e.g. crystals of CO) and models for the glassy state (Ziman, 1979 ; Stillinger and Weber, 1982 ; Stillinger and Weber, 1984 ; Toulouse, 1984 ; Ansari et al., 1985 ; Stein, 1985). A number of experiments suggest that the transition temperature for myoglobin is in the neighborhood of 200°K (Austin et al., 1975 ; Parak et al., 1982 ; Debrunner and Frauenfelder, 1962 ; Ansari et al., 1985). Because large scale, collective motions that involve the protein surface are important in the fluctuations (Swaminathan et al., 1982), it is likely that the observed transition is due to the freezing of the solvent matrix (Swaminathan et al., 1982 ; Parak et al., 1982).

Because the details of the native structure of a protein play an essential role in its function, it is important to determine the structural origins of the

multiminimum surface obtained from the dynamics analysis. The general features of the structure (helices and turns) are preserved throughout the simulation and the differences in position are widely distributed. The motions are associated primarily with loop displacements or relative displacements of helices which individually behave as nearly rigid bodies. Rearrangements within individual loops are the elementary step in the transition from one minimum to another ; they are coupled with associated helix displacements. Which loop or turn changes in a given time interval appears to be random. Specific loop motions may be initiated by sidechain transitions in the helix contacts, mainchain dihedral angle transitions of the loops themselves, or a combination of the two. As the time interval between two structures increases, more loop transitions have occurred. At room temperature, the transition probabilities are such that for an interval 100 psec or longer between two structures, some transitions will have taken place in all of the flexible loop regions. However, since the rms differences between structures continue to increase up to 200 psec, the configuration space available to the molecule includes a range of structures for the loop regions that are not completely sampled in a 100 psec.

To characterize the helix motions that are coupled with the loop rearrangements, the internal structural changes of the helices were separated from their relative motions. Individual helices and loops were superimposed and the rms differences for the mainchain calculated for the set of structures ; the rms difference for the internal structure of the helices is generally less than 1 Å. Corresponding results for the loop regions show that they undergo much larger internal structural changes on the order of 2.5 Å.

In analysing the relative motion of the helices, it is of particular interest to examine the behavior of helix pairs that are in van der Waals contact ; these are helix pairs A-H, B-E, B-G, F-H and G-H for all of which at least three residues from each helix are interacting. Each helix was fitted to a straight line and the fluctuations of the distance between the helix centers of mass and the relative orientations of the lines were compared. The relative translations found in this case have rms values of 0.3 to 0.7 Å and the relative rotations have rms values of 1 to 14°; the maximum differences are 1.3 to 2.2 Å and 5 to 39°, respectively.

The dynamical results for the helix motions can be compared with structural data from two sources ; the first is derived from proteins of a given sequence in different environments (e.g., two different crystal forms, deoxy and oxy hemoglobin ; Chothia and Lesk, 1985) and the second from homologous proteins with different sequences (e.g., the globins ; Lesk and Chothia, 1980). The maximum dynamical displacements are, in fact, larger than those observed in different x-ray structures of a given protein. The values are of the same order as the differences (2 to 3 Å, 15 to 30° ; there are some larger changes) found in comparing a series of different globins with known crystal structures and

sequence homology in the range 16 to 88 percent. Thus, the range of conformations sampled by a single myoglobin trajectory is similar to that found in the evolutionary variation among crystal structures of the globin series. This suggests a molecular plasticity which is likely to have played an important role in the evolution of protein sequences.

The comparison of the various globin structures (Lesk and Chothia, 1980) suggested that the range of helix packings is achieved primarily by changes in sidechain volumes resulting from amino acid substitutions. In the dynamics, it is the correlated motions of sidechains that are in contact, plus the rearrangements of loops, that make possible the observed helix fluctuations. Different positions within wells and transitions between wells for sidechains (e.g., $\pm 60^\circ$, 180° for X_1) are involved. This is in accord with the results of high-resolution x-ray studies that show significant disorder in sidechain orientations (Smith et al., 1986; Kuriyan et al., 1987). Further, correlated dihedral angle changes differentiate the various minima. Since more than one set of sidechain orientations is consistent with a given set of helix positions, the known globin crystal structures probably represent only a small subset of the possible local minima.

Myoglobin at normal room temperatures samples a very large number of different minima that arise from the inhomogeneity of the system. This is expected to have important consequences for the interpretation of myoglobin function and, more generally, for the functions of other proteins, including enzymes. There are solid-like microdomains (the helices), whose mainchain structure is relatively rigid, and liquid-like regions (the loops and the sidechain clusters at interhelix contacts) that readjust as the helices move from one minimum to another. Since the minima have similar energies myoglobin is expected to be glass-like at low temperatures. Freezing in of the liquid-like regions could result in a transition to the glassy state (Stein, 1985).

ATOMIC FLUCTUATION AND X-RAY DIFFRACTION

Since atomic fluctuations are the basis of protein dynamics, it is important to have experimental tests of the accuracy of the simulation results concerning them. For the magnitudes of the motions, the most detailed data are provided, in principle, by an analysis of the Debye-Waller or temperature factors obtained in crystallographic refinements of x-ray structures. Averages over the fluctuations can be obtained also from incoherent neutron scattering (Doster et al., 1989).

It is well known from small-molecule crystallography that the effects of thermal motion must be included in the interpretation of the x-ray data to obtain accurate structural results. Detailed models have been introduced to take account of anisotropic and anharmonic options of the atoms and these molecules have been applied to high-resolution data for small molecules (Zucker and Schultz, 1982). In protein crystallography, the limited data available relative to the large

number of parameters that have to be determined, have made it necessary in most cases to assume that the atomic motions are isotropic and harmonic. Then the structure factor, $F(\underline{Q})$, which is related to the measured intensity by $I(\underline{Q}) = |F(\underline{Q})|^2$, is given by

$$F(\underline{Q}) = \sum_{j=1}^N f_j(\underline{Q}) e^{i\underline{Q} \cdot \langle \underline{r}_j \rangle} e^{W_j(\underline{Q})} \quad (3)$$

where \underline{Q} is the scattering vector, $\langle \underline{r}_j \rangle$ is the average position of atom j with atomic scattering factor $f_j(\underline{Q})$ and the sum is over the N atoms in the asymmetric unit of the crystal. The Debye-Waller factor, $W_j(\underline{Q})$, is defined by

$$W_j(\underline{Q}) = -\frac{8}{3} \pi^2 \langle \Delta r_j^2 \rangle s^2 = -B_j s^2 \quad (4)$$

where $s = |\underline{Q}|/4\pi$. The quantity B_j is usually referred to as the temperature factor, which is directly related to the mean-square atomic fluctuations in the isotropic harmonic model. More generally, if the motion is harmonic but anisotropic, a set of six parameters

$$B_j^{xx} = \langle \Delta x_j^2 \rangle, B_j^{xy} = \langle \Delta x_j \Delta y_j \rangle, \dots, B_j^z = \langle \Delta z_j^2 \rangle \quad (5)$$

is required to fully characterize the atomic motion. Although in the earlier x-ray studies of proteins, the significance of the temperature factors was ignored (presumably because the data were not at a sufficient level of resolution and accuracy), more recently attempts have been made to relate the observed temperature factors to the atomic motions (Frauenfelder et al., 1979 ; Artymiuk et al., 1979).

In principle, the temperature factors provide a very detailed measure of these motions because information is available for the mean-square fluctuation of every heavy atom. In practice, there are two types of difficulties in relating the B factors obtained from protein refinements to the atomic motions. The first is that, in addition to thermal fluctuations, any static (lattice) disorder in the crystal contributes to the B factors ; i.e., since a crystal is made up of many unit cells, different molecular geometries in the various cells have the same effect on the average electron density, and therefore the B factor, as atomic motions. For the iron atom of myoglobin there has been an experimental attempt to determine the disorder contribution (Hartmann et al., 1982). Since the Mossbauer effect is not altered by static disorder (i.e., each nucleus absorbs independently), but does depend on atomic motions, comparisons of Mossbauer and x-ray data have been

used to estimate a disorder contribution for the iron atom ; the value obtained is

$$\langle \Delta r_{\text{Fe}}^2 \rangle = 0.08 \text{ \AA}^2$$

Although the value is only approximate, it nevertheless indicates that the observed B factors (e.g., on the order of 0.44 \AA^2 for backbone atoms and 0.50 \AA^2 for sidechain atoms) are dominated by the motional contribution. Most experimental B factor values are compared directly with the molecular dynamics results (i.e., neglecting the disorder contribution) or are rescaled by a constant amount (e.g., by setting the smallest observed B factor to zero) on the assumption that the disorder contribution is the same for all atoms (Petsko and Ringe, 1984). The second difficulty is that, since simulations have shown that the atomic fluctuations are highly anisotropic and, in some cases, anharmonic, there may be significant errors in the refinement due to the assumption of isotropic and harmonic motion. A direct experimental estimate of the errors is difficult because sufficient data are not yet available for protein crystals, although incoherent neutron scattering can provide information independent of static disorder (Doster et al., 1989). Moreover, any data set includes other errors which would obscure the analysis. As an alternative to an experimental analysis of the errors in the refinement of proteins, a purely theoretical approach can be used (Kuriyan et al., 1986). The basic idea is to generate x-ray data from a molecular dynamics simulation of a protein and to use these data in a standard refinement procedure. The error in the analysis can then be determined by comparing the refined x-ray structure and temperature factors with the average structure and the mean-square fluctuations from the simulation. Such a comparison, in which no real experimental results are used, avoids problems due to inaccuracies in the measured data (exact calculated intensities are used), to crystal disorder (there is none in the model), and to approximations in the simulation (the simulation is exact for this case). The only question about such a comparison is whether the atomic motions found in the simulation are a meaningful representation of those occurring in proteins. As has been shown (Petsko and Ringe, 1984 ; Karplus and McCammon, 1983), molecular dynamics provides a reasonable picture of the motions in spite of errors in the potentials, the neglect of the crystal environment, and the finite-time classical trajectories used to obtain the results. However, these inaccuracies do not affect the exactitude of the computer "experiment" for testing the refinement procedure that is described below.

In this study (Kuriyan et al., 1986), a 25- μ sec molecular dynamics trajectory for myoglobin was used (Levy et al., 1985). The average structure and the mean-square fluctuations from that structure were calculated directly from the trajectory. To obtain the average electron density, appropriate atomic electron distributions were assigned to the individual atoms and the results for each

coordinate set were averaged over the trajectory. Given the symmetry, unit cell dimensions, and position of the myoglobin molecule in the unit cell, average structure factors, $\langle F(Q) \rangle$, and intensities, $I(Q) = |F(Q)|^2$, were calculated from the Fourier transform of the average electron density, $\langle \rho(r) \rangle$, as a function of position r in the unit cell. Data were generated at 1.5 Å resolution, since this is comparable to the resolution of the best x-ray data currently available for proteins the size of myoglobin (Kuriyan et al., 1987). The resulting intensities at Bragg reciprocal lattice points were used as input data for the widely applied crystallographic program, PROLSQ (Konnert and Hendrickson, 1980). The time-averaged atomic positions, obtained from the simulation and a uniform temperature factor provide the initial model for refinement. The positions and an isotropic, harmonic temperature factor for each atom were then refined iteratively against the computer-generated intensities in the standard way. Differences between the refined results for the average atomic positions and their mean-square fluctuations and those obtained from the molecular dynamics trajectory are due to errors introduced by the refinement procedure.

The overall rms error in atomic positions ranged from 0.24 Å to 0.29 Å for slightly different restrained and unrestrained refinement procedures (Kuriyan et al., 1986). The errors in backbone positions (0.10 - 0.20 Å) are generally less than those for sidechain atoms (0.28 - 0.33 Å); the largest positional errors are on the order of 0.6 Å. The backbone errors, although small, are comparable to the rms deviation of 0.21 Å between the positions of the backbone atoms in the refined experimental structures of oxymyoglobin and carboxymyoglobin (Kuriyan et al., 1987; Phillips, 1980). Further, the positional errors are not uniform over the whole structure. There is a strong correlation between the positional error and the magnitude of the mean-square fluctuation for an atom, with certain regions of the protein, such as loops and external sidechains, having the largest errors.

The refined mean-square fluctuations are systematically smaller than the fluctuations calculated directly from the simulation. The magnitudes and variation of temperature factors along the backbone are relatively well reproduced, but the refined sidechain fluctuations are almost always significantly smaller than the actual values. The average backbone B factors from different refinements are in the range 11.3 to 11.7 Å², as compared with the exact value of 12.4 Å²; for the sidechains, the refinements yield 16.5 to 17.6 Å², relative to the exact value of 26.8 Å². Regions of the protein that have high mobility have large errors in temperature factors as well as in positions. Examination of all atoms shows that fluctuations greater than about 0.75 Å² ($B = 20$ Å²) are almost always underestimated by the refinement. Moreover, while actual mean-square atomic fluctuations have values as large as 5 Å², the x-ray refinement leads to an effective upper limit of about 2 Å². This arises from the fact that most of the

atoms with large fluctuations have multiple conformations and that the refinement procedure picks out one of them.

To do refinements that take some account of anisotropic motions for all but the smallest proteins, it has been necessary to introduce assumptions concerning the nature of the anisotropy. One possibility is to assume anisotropic rigid body motions for sidechains such as tryptophan and phenylalanine (Artymiuk et al., 1979; Glover et al., 1983). An alternative is to introduce a "dictionary" in which the orientation of the anisotropy tensor is related to the stereochemistry around each atom (Konnert and Hendrickson, 1980); this reduces the six independent parameters of the anisotropic temperature factor tensor B_j to three parameters per atom. An analysis of a simulation for BPTI (Yu et al., 1985) has shown that the actual anisotropies in the atomic motions are generally not simply related to the local stereochemistry; an exception is the mainchain carbonyl oxygen, which has its largest motion perpendicular to the C=O bond. Thus, use of stereochemical assumptions in the refinement can yield incorrectly oriented anisotropy tensors and significantly reduced values for the anisotropies. The large-scale motions of atoms are collective and sidechains tend to move as a unit so that the directions of largest motion are not related to the local bond direction and have similar orientations in the different atoms forming a group that is undergoing correlated motions. Consequently, it is necessary to use the full anisotropy tensor to obtain meaningful results. This is possible with proteins that are particularly well ordered, so that the diffraction data extend to better than 1 Å resolution.

X-RAY REFINEMENT BY SIMULATED ANNEALING

Crystallographic structure determinations by x-ray or neutron diffraction generally proceed in two stages. First, the phases of the measured reflections are estimated and a low- to medium-resolution model of the protein is constructed and second, more precise information about the structure is obtained by refining the parameters of the molecular model against the crystallographic data (Wyckoff et al., 1985). The refinement is performed by minimizing the crystallographic R factor, which is defined as the difference between the observed ($|F_{\text{obs}}(h,k,l)|$) and calculated ($|F_{\text{calc}}(h,k,l)|$) structure factor amplitudes,

$$R = \sum_{h,k,l} \left| |F_{\text{obs}}(h,k,l)| - |F_{\text{calc}}(h,k,l)| \right| / \sum_{h,k,l} |F_{\text{obs}}(h,k,l)| \quad (6)$$

where h,k,l are the reciprocal lattice points of the crystal.

Conventional refinement involves a series of steps, each consisting of a few cycles of least-squares refinement with stereochemical and internal packing constraints or restraints (Sussman et al., 1977; Jack and Levitt, 1978; Konnert

and Hendrickson, 1980 ; Moss and Morffew, 1982) that are followed by manual rebuilding of the model structure by use of interactive computer graphics. Finally, solvent molecules are included and alternative conformations for some protein atoms may be introduced. The standard refinement procedure is time consuming, because the limited radius of convergence of least-squares algorithms (approximately 1 Å) necessitates the periodic examination of electron density maps computed with various combinations of F_{obs} and F_{calc} as amplitudes, and with phases calculated from the model structure. Also, the least-squares refinement process is easily trapped in a local minimum so that human intervention is necessary.

Simulated annealing (Kirkpatrick et al., 1983), which makes use of Monte Carlo or molecular dynamics (Brünger et al., 1987a) simulations to explore the conformational space of the molecule can help to overcome the local-minimum problem. This has been demonstrated in the application of molecular dynamics to structure determination with nuclear magnetic resonance (NMR) data. In contrast to the NMR application (Brünger et al., 1986), the initial model for crystallographic refinement cannot be arbitrary. It has to be relatively close to the correct geometry to provide an adequate approximation to the phases of the structure factors.

To employ molecular dynamics in crystallographic refinement, an effective potential

$$E_{sf} = S \sum_{h,k,l} \left[\left| F_{obs}(h,k,l) \right| - \left| F_{calc}(h,k,l) \right| \right]^2 \quad (7)$$

was added to the empirical energy potential given in Eq. 1. The effective potential E_{sf} describes the differences between the observed structure factor amplitudes and those calculated from the atomic model ; it is identical to the function used in standard least-squares refinement methods (Jack and Levitt, 1978). The scale factor S was chosen to make the gradient of E_{sf} comparable in magnitude to the gradient of the empirical energy potential of a molecular dynamics simulation with S set to zero.

As in the case of the NMR analysis, simulated annealing refinement was also tested on crambin, for which high-resolution x-ray diffraction data and a refined structure, determined by resolved anomalous phasing and conventional least-squares refinement with model-building, are available (Hendrickson and Teeter, 1981). The initial structure for the MD-refinement was obtained from the NMR structure determination (see Sect. IV) ; the orientation and position of the NMR-derived crambin molecule in the unit cell was determined by molecular replacement (Brünger et al., 1987b). The root-mean-square (rms) differences for residue positions of this initial structure and the final manually refined structure

(Hendrickson and Teeter, 1981) are as large as 3.5 Å, with particularly large differences for residues 34 to 40 ; the R factor of the initial structure was 0.56 at 2 Å resolution. MD-refinement at 3000 K starting with 4 Å resolution data for 2.5 ps, extending to 3 Å resolution for 2.5 ps, and finally to 2 Å resolution for 5 ps, followed by several cycles of minimization, reduces the atomic rms deviations to 0.34 and 0.56 Å for the backbone and sidechain atoms, respectively. During the MD-refinement, some atoms in residues 35 to 40 moved by more than 3 Å. The essential point is that the refinement of the crambin structure was achieved starting from the initial NMR-structure without human intervention. The R factor (0.294) of the MD-refined structure is somewhat higher than the R factor (0.258) of the manually refined structure without solvent and with constant temperature factors ; minor model-building would correct this difference. Other annealing protocols using higher temperatures (e.g., 7000 to 9000°K) yield structures that are still closer to the manually refined structure. The refinement required approximately one hour of central processing unit (CPU) time on CRAY-1 ; structure factor calculations accounted for about half this time. The latter portion of the calculation has been considerably reduced in time by use of Fast Fourier Transform (FFT) methods (Brünger, 1989).

As a control, the initial NMR-derived structure was refined without rebuilding by a restrained least-squares method (Konnert and Hendrickson, 1980), starting at 4 Å resolution and then increasing the resolution to 3 Å, and finally to 2 Å. The R factor dropped to 0.381, but the very bad stereochemistry and large deviation from the manually refined structure indicate that this structure has not converged to the correct result ; residues 34 to 40 have not moved and substantial model-building would be required to correct the structure. Thus, restrained least-squares refinement in the absence of model-building did not produce the large conformational changes that occurred in MD-refinement by simulated annealing. With a version of the molecular dynamics program CHARMM optimized for x-ray refinement (the program X-PLOR, Brünger et al., 1989), many applications of simulated annealing have been made and shown to be of considerable utility in decreasing the human effort involved (e.g., Navia et al., 1989).

USE OF NUCLEAR MAGNETIC RESONANCE DATA FOR DYNAMICS AND STRUCTURE

Nuclear magnetic resonance (NMR) is an experimental technique that has played an essential role in the analysis of the internal motions of proteins (Campbell et al., 1978 ; Gurd and Rothgeb, 1979 ; Dobson and Karplus, 1986). Like x-ray diffraction, it can provide information about individual atoms ; unlike x-ray diffraction, NMR is sensitive not only to the magnitude but also to the time scales of the motions. Most nuclear relaxation processes are dependent on atomic

motions on the nanosecond to picosecond time scale. Although molecular tumbling is generally the dominant relaxation mechanism for proteins in solution, internal motions contribute as well ; for solids, the internal motions are of primary importance. In addition, NMR parameters, such as nuclear spin-spin coupling constants and chemical shifts, depend on the protein environment. In many cases different local conformations exist but the interconversion is rapid on the NMR time scale, here on the order of milliseconds, so that average values are observed. When the interconversion time is on the order of the NMR time scale or slower, the transition rates can be studied by NMR ; an example is provided by the reorientation of aromatic rings (Campbell et al., 1976 ; Brooks et al., 1987).

In addition to supplying data on the dynamics of proteins, NMR can also be used to obtain structural information. With recent advances in techniques it is now possible to obtain a large number of approximate interproton distances for proteins by the use of nuclear Overhauser effect measurements (Noggle and Schirmer, 1971). If the protein is relatively small and has a well resolved spectrum, a large portion of the protons can be assigned and several hundred distances for these protons can be determined by the use of two-dimensional NMR techniques (Wagner and Wüthrich, 1982). Clearly, these distances can serve to provide structural information for proteins, analogous to their earlier use for organic molecules (Noggles and Schirmer, 1971 ; Honig et al., 1971). Of great interest is the demonstration that enough distance information can be measured to determine the high resolution structure of a protein in solution. In the last few years it has been shown how such NMR structures can serve to supplement results from x-ray crystallography, particularly for proteins that are difficult to crystallize (Wüthrich, 1989).

In what follows we consider two questions related to structure determination. The first concerns the effect of motional averaging on the accuracy of the apparent distances obtained from the NOE studies and the second, the use of molecular dynamics simulated annealing to obtain structural results from the NOE data.

For spin-lattice relaxation, such as observed in nuclear Overhauser effect measurements, it is possible to express the behavior of the magnetization of the nuclei being studied by the equation (Olejniczak et al., 1984 ; Solomon, 1955).

$$\frac{d(I_z(t) - I_0)_i}{dt} = -\rho_i (I_z(t) - I_0)_i - \sum_{i \neq j} \sigma_{ij} (I_z(t) - I_0)_j \quad (8)$$

where $I_z(t)_i$ and I_{0i} are the z components of the magnetization of nucleus i at time t and at equilibrium, ρ_i is the direct relaxation rate of nucleus i, and σ_{ij} is the cross relaxation rate between nuclei i and j. The quantities ρ_i and σ_{ij} can be

expressed in terms of spectral densities

$$\begin{aligned}\rho_i &= \frac{6\pi}{5} \gamma_i^2 \gamma_j^2 \hbar^2 \sum_{i \neq j} \left[\frac{1}{3} J_{ij}(\omega_i - \omega_j) + J_{ij}(\omega_i) + 2J_{ij}(\omega_i - \omega_j) \right] \\ \sigma_{ij} &= \frac{6\pi}{5} \gamma_i^2 \gamma_j^2 \hbar^2 \left[2J_{ij}(\omega_i - \omega_j) - \frac{1}{3} J_{ij}(\omega_i - \omega_j) \right]\end{aligned}\quad (9)$$

where ω_i is the resonance frequency of nucleus i . The spectra density functions can be obtained from the correlation functions for the relative motions of the nuclei with spins i and j (Olejniczak et al., 1984 ; Levy et al., 1981),

$$J_{ij}^n(\omega) = \int_0^\infty \frac{\langle Y_n^2(\Theta_{lab}(t)\phi_{lab}(t)) Y_n^{2*}(\Theta_{lab}(0)\phi_{lab}(0)) \rangle}{r_{ij}^3(0)r_{ij}^3(t)} \cos(\omega t) dt \quad (10)$$

where $Y_n^2(\Theta(t)\phi(t))$ are second-order spherical harmonics and the angular brackets represent an ensemble average which is approximated by an integral over the molecular dynamics trajectory. The quantities $\Theta_{lab}(t)$ and $\phi_{lab}(t)$ are the polar angles at time t of the internuclear vector between protons i and j with respect to the external magnetic field and r_{ij} is the interproton distance. In the simplest case of a rigid molecule undergoing isotropic tumbling with a correlation time τ_0 this reduces to the familiar expression

$$J_{ij}(\omega) = \frac{1}{4\pi r_{ij}^6} \left[\frac{\tau_0}{1 + (\omega\tau_0)^2} \right] \quad (11)$$

The nuclear Overhauser effect corresponds to the selective enhancement of a given resonance by the irradiation of another resonance in a dipolar coupled spin system. Of particular interest for obtaining motional and distance information are measurements that provide time-dependent NOEs from which the cross relaxation rates σ_{ij} (see Eq. 9) can be determined directly or indirectly by solving a set of coupled equations (Eqs. 8 and 9). Motions on the picosecond timescale are expected to introduce averaging effects that decrease the cross-relaxation rates by a scale factor relative to the rigid model. A lysozyme molecular dynamics simulation (Ichiye et al., 1986) has been used to calculate dipole vector correction functions (Olejniczak et al., 1984) for proton pairs that have been studied experimentally (Olejniczak et al., 1981 ; Poulsen et al., 1980). Four proton pairs on three sidechains (Trp 28, Ile 98 and Met 105) with very different motional properties were examined. Trp 28 is quite rigid, Ile 98 has significant fluctuations, and Met 105 is particularly mobile in that it jumps among

different side-chain conformations during the simulation. The rank order of the scale factors (order parameters) is the same in the theoretical and experimental results. However, although the results for the Trp 28 protons agree with the measurements to within the experimental error, for both Ile 98 and Met 105 the motional averaging found from the NOE's is significantly greater than the calculated value. This suggests that these residues are undergoing rare fluctuations involving transitions that are not adequately sampled by the simulation.

If nuclear Overhauser effects are measured between pairs of protons whose distance is not fixed by the structure of a residue, the strong distance-dependence of the cross-relaxation rates ($1/r^6$) can be used to obtain estimates of the interproton distances (Poulsen et al., 1980 ; Olejniczak et al., 1981 ; Wagner and Wüthrich, 1982 ; Clore et al., 1985). The simplest application of this approach is to assume that proteins are rigid and tumble isotropically. The lysozyme molecular dynamics simulation was used to determine whether picosecond fluctuations are likely to introduce important errors into such an analysis (Olejniczak et al., 1984). The result shows that the presence of the motions will cause a general decrease in most NOE effects observed in a protein. However, because the distance depends on the sixth root of the observed NOE, motional errors of a factor of two in the latter lead to only a 12% uncertainty in the distance. Thus, the decrease is usually too small to produce a significant change in the distance estimated from the measured NOE value. This is consistent with the excellent correlation found between experimental NOE values and those calculated using distances from a crystal structure (Poulsen et al., 1980). Specific NOEs can, however, be altered by the internal motions to such a degree that the effective distances obtained are considerably different from those predicted for a static structure. Such possibilities must, therefore, be considered in any structure determination based on NOE data. This is true particularly for cases involving averaging over large fluctuations, such as may occur for external sidechains and mobile loop regions.

Because of the inverse sixth power of the NOE distance dependence experimental data so far are limited to protons that are separated by less than 5 Å. Thus, the information required for a direct protein structure determination is not available. To overcome this limitation it is possible to introduce additional information provided by empirical energy functions (Brook et al., 1983). One way of proceeding is to do molecular dynamics simulated annealing with the approximate interproton distances introduced as restraints in the form of skewed biharmonic potentials (Clore et al., 1985 ; Brünger et al., 1986) ; the force constants can be chosen to correspond to the experimental uncertainty in the distance.

A model study of the small protein crambin, which is composed of 46 residues, was made with realistic NOE restraints (Brünger et al., 1986). Two

hundred forty approximate interproton distances less than 4 Å were used, including 184 short-range distances (i.e., those connecting protons in two residues that are less than 5 residues apart in the sequence) and 56 long-range distances. The molecular dynamics simulations converged to the known crambin structure (Hendrickson and Teeter, 1981) from different initial extended structures. The average structure obtained from the simulations with a series of different protocols had rms deviations of 1.3 Å for the backbone atoms, and 1.9 Å for the sidechain atoms. Individual converged simulations had rms deviations in the range 1.5 to 2.1 Å and 2.1 to 2.8 Å for the backbone and sidechain atoms, respectively. Further, it was shown that a dynamics structure with significantly large deviations (5.7 Å) could be characterized as incorrect, independent of a knowledge of the crystal structure because of its higher energy and the fact that the NOE restraints were not satisfied within the limits of error. The incorrect structure resulted when all NOE restraints were introduced simultaneously, rather than allowing the dynamics to proceed first in the presence of only the short-range restraints followed by introduction of the long-range restraints. Also of interest is the fact that although crambin has three disulfide bridges it was not necessary to introduce information concerning them to obtain an accurate structure.

The folding process as simulated by the restrained dynamics is very rapid. At the end of the first 2 ps the secondary structure is essentially established while the molecule is still in an extended conformation. Some tertiary folding occurs even in the absence of long-range restraints. When they are introduced, it takes about 5 ps to obtain a tertiary structure that is approximately correct and another 6 ps to introduce the small adjustments required to converge to the final structure.

It is of interest to consider the relation between the results obtained in the restrained dynamics simulation and actual protein folding. That correctly folded structures are achieved only when the secondary structural elements are at least partly formed before the tertiary restraints are introduced is suggestive of the diffusion-collision model of protein folding (Bashford et al., 1984). Clearly, the specific pathway has no physical meaning since it is dominated by the NOE restraints. Also, the time scale of the simulated folding process is 12 orders of magnitude faster than experimental estimates. About 6 to 9 orders of magnitude of the rate increase are due to the fact that the secondary structure is stable once it is formed, in contrast to real protein where the secondary structural elements spend only a small fraction of time in the native conformation until coalescence has occurred. The remainder of the artificial rate increase presumably arises from the fact that the protein follows a single fairly direct path to the folded state in the presence of the NOE restraints, instead of having to go through a complex search process.

Many applications of NMR data to structure determinations have been made. Both distance geometry methods and molecular dynamics have been

employed for reducing the data (Wüthrich, 1989 ; Clore and Gronenborn, 1989).

STRUCTURAL ROLE OF ACTIVE-SITE WATERS IN RIBONUCLEASE A

To achieve a realistic treatment of solvent-accessible active sites, a molecular dynamics simulation method, called the stochastic boundary method, has been implemented (Brooks and Karplus, 1983 ; Brünger et al., 1984 ; Brooks et al., 1985). It makes possible the simulation of a localized region, approximately spherical in shape, that is composed of the active site with or without ligands, the essential portions of the protein in the neighborhood of the active site, and the surrounding solvent. The approach provides a simple and convenient method for reducing the total number of atoms included in the simulation, while avoiding spurious edge effects.

The stochastic boundary method for solvated proteins starts with a known x-ray structure ; for the present problem the refined high-resolution (1.5 to 2 Å) x-ray structures provided by Petsko and coworkers was used (G. Petsko, private communication). The region of interest (here the active site of ribonuclease A) was defined by choosing a reference point (which was taken at the position of the phosphorus atom in the CpA inhibitor complex) and constructing a sphere of 12 Å radius around this point. Space within the sphere not occupied by crystallographically determined atoms was filled by water molecules, introduced from an equilibrated sample of liquid water. The 12-Å sphere was further subdivided into a reaction region (10 Å radius) treated by full molecular dynamics and a buffer region (the volume between 10 and 12 Å) treated by Langevin dynamics, in which Newton's equations of motion for the nonhydrogen atoms are augmented by a frictional term and a random-force term ; these additional terms approximate the effects of the neglected parts of the system and permit energy transfer in and out of the reaction region. Water molecules diffuse freely between the reaction and buffer regions, but are prevented from escaping by an average boundary force (Brünger et al., 1984). The protein atoms in the buffer region are constrained by harmonic force derived from crystallographic temperature factors (Brooks et al., 1985). The forces on the atoms and their dynamics were calculated with the CHARMM program (Brooks et al., 1983) ; the water molecules were represented by the ST2 model (Stillinger and Rahman, 1974).

One of the striking aspects of the active site of ribonuclease is the presence of a large number of positively charged groups, some of which may be involved in guiding and/or binding the substrate (Matthew and Richards, 1982). The simulation demonstrated that these residues are stabilized in the absence of ligands by well-defined water networks. A particular example includes Lys-7, Lys-41, Lys-66, Arg-39 and the doubly protonated His-119. Bridging waters, some of

which are organized into trigonal bipyramidal structures, were found to stabilize the otherwise very unfavorable configuration of near-neighbor positive groups because the interaction energy between water and the charged C-NH^+_n ($n = 1, 2$, or 3) moieties is very large ; e.g., at a donor-acceptor distance of 2.8 \AA , the $\text{C-NH}^+_3 - \text{H}_2\text{O}$ energy is -19 kcal/mole with the empirical potential used for the simulation (Brooks et al., 1983), in approximate agreement with accurate quantum-mechanical calculations (Desmeules and Allen, 1980) and gas-phase ion-molecule data (Kearle, 1977). The average stabilization energy of the charged groups (Lys-7, Lys-41, Lys-66, Arg-39 and His-119) and the 106 water molecules included in the simulation is -376.6 kcal/mole . This energy is calculated as the difference between the simulated system and a system composed of separate protein and bulk water. Unfavorable protein-protein charged-group interactions are balanced by favorable water-protein and water-water interactions. The average energy per molecule of pure water from an equivalent stochastic boundary simulation (Brünger et al., 1984) was -9.0 kcal/mole , whereas that of the waters included in the active-site simulation was -10.2 kcal/mole ; in the latter a large contribution to the energy came from the interactions between the water molecules and the protein atoms. It is such energy differences that are essential to a correct evaluation of binding equilibria and the changes introduced by site-specific mutagenesis (Fersht et al., 1985).

During the simulation, the water molecules involved in the charged-group interactions oscillated around their average positions, generally without performing exchange. On a longer time scale, it is expected that the waters would exchange and that the sidechains would undergo larger scale displacements. This is in accord with the disorder found in the x-ray results for lysine and arginine residues (e.g., Lys-41 and Arg-39) (Gilbert et al., to be published ; Wlodawer, 1985), a fact that makes difficult a crystallographic determination of the water structure in this case. It is also of interest that Lys-7 and Lys-41 have an average separation of only 4 \AA in the simulation, less than that found in the x-ray structure. That this like-charged pair can exist in such a configuration is corroborated by experiments that have shown that the two lysines can be cross-linked (Marfey et al., 1965) ; the structure of this compound has been reported recently (Weber et al., 1985) and is similar to that found in the native protein.

In addition to the role of water in stabilizing the charged groups that span the active site and participate in catalysis, water molecules make hydrogen bonds to protein polar groups that become involved in ligand binding. A particularly clear example is provided by the adenine-binding site in the CpA simulation. The NH_2 group of adenine acted as a donor, making hydrogen bonds to the carbonyl of Asn-67, and the ring N1A of adenine acted as an acceptor for a hydrogen bond from the amide group of Glu-69. Corresponding hydrogen bonds were present in

the free ribonuclease simulation, with appropriately bound water molecules replacing the substrate. These waters and those that interact with the pyrimidine-site residues Thr-45 and Ser-123 help to preserve the protein structure in the optimal arrangement for binding. Similar substrate "mimicry" has been observed in x-ray structures of lysozyme (Blake et al., 1983) and of penicillopepsin (James and Sielecki, 1983), but has not yet been seen in ribonuclease.

A COOPERATIVITY MUTANT IN HEMOGLOBIN

Hemoglobin has long been a subject of experimental and theoretical studies because it is the classic example of cooperativity in biological systems. Since the determination of the x-ray structure of the unliganded (deoxy) and liganded (oxy) tetramers by Perutz and coworkers (Perutz, 1970 ; Fermi and Perutz, 1981), attention has been focused on the atomic details of the cooperative mechanism. In particular, structural and thermodynamic measurements on native, mutant, and modified hemoglobins have been utilized in attempts to isolate the essential amino acids and to determine their contributions to cooperativity. It has been suggested on the basis of such studies (Pettigrew et al., 1982 ; Perutz, 1970) and theoretical analyses (Gelin et al., 1983) that the interactions between the C helix of one chain and the FG corner region of another ($C\alpha_1$ - FGB_2) play an important role in the coupling of relative stabilities of the quaternary structures of the tetramer to the tertiary changes induced in the allosteric core by ligand binding to individual subunits. Asp $\beta 99$ (G1) is one of the residues that have been studied in great detail. A series of naturally occurring mutants all have significantly reduced cooperativity and increased oxygen affinity relative to normal hemoglobin (Dickerson and Geis, 1983 ; Bunn and Forget, 1980). From a comparison of the deoxy and oxy normal hemoglobin tetramer structures, it has been suggested that the essential role of Asp $\beta 99$ (G1) is to stabilize the deoxy tetramer by making hydrogen bonds to Tyr $\alpha 42$ (C7) and to Asn $\alpha 97$ (G4), which are absent in the oxy tetramer (Morimoto et al., 1971). It is now possible to supplement such observational conclusions by free energy simulations. We here employ the simulation method to show that the observed free energy changes result from interactions of Asp $\beta 99$ (G1) with several amino acids and with the solvent. Although both Tyr $\alpha 42$ (C7) and Asn $\alpha 97$ (G4) are found to play a significant role, other interactions are shown to be of equal or greater importance. In what follows, we focus on the mutant Asn $\beta 99$ (G1) to Ala (Hb Radcliffe) (Weatherall et al., 1977) which is of the "deletion" type (Ferscht, 1987) and is therefore expected to lead to only localized structural changes (Shih et al., 1985) that are simplest to interpret.

The free energy difference ΔG between two states A and B (here A corresponds to normal hemoglobin and B to a mutant hemoglobin) is obtained by

thermodynamic integration with the formula (Kirkwood, 1935 ; Kirkwood, 1942 ; Fleischman et al., 1990).

$$\Delta G = \int_0^1 \langle \Delta V \rangle_\lambda d\lambda \quad (12)$$

where $\Delta V = V_B - V_A$ and λ is a parameter, such that $V_\lambda = (1-\lambda) V_A + \lambda V_B$; the quantities V_A and V_B are empirical energy functions describing, respectively, the normal and the mutant hemoglobin molecule system. The essential part of the calculation is the evaluation of the thermodynamic average $\langle \Delta V \rangle_\lambda$, where the subscript λ implies that average is over the hybrid system described by V_λ . For calculating the integral (Eq. 12), a series of λ values are used (Fleischman et al., 1990 ; Brooks et al., 1983). A stochastic boundary simulation, which followed the procedure described previously (Brünger et al., 1985), was employed to obtain the required averages.

To determine the effect of the mutation Asp $\beta 99$ (G1) \rightarrow Ala on cooperativity, the free energy change of the deoxy and oxy tetramers resulting from this mutation has been calculated by Eq. (12) (Gao et al., 1989) (see Table 1). Both the deoxy and the oxy tetramer are destabilized by the mutation (66 and 60.5 kcal/mole, respectively per interface), but it is the deoxy tetramer which is more destabilized, leading to the reduced cooperativity and increased ligand affinity. The differences between them (5.5 kcal/mole) can be compared with the measurements of Ackers et al. (3.4 kcal/mole ; private communication). The experimental and theoretical results have the same sign and are of the same order, suggesting that the simulation may be meaningfully analyzed to obtain insight into the interactions that contribute to the free energy differences.

To analyze the results, we make use of the fact that due to the linear form of Eq. (12), the free energy can be decomposed into the contribution from interactions between the mutated residue and any other residues or water molecules ; in all cases, we consider the change in the free energy induced by the mutation. The change in the solvent interactions, which are essentially electrostatic, is more destabilizing for the oxy than for the deoxy tetramer. This is in accord with the x-ray structures since in the oxy tetramer the Asp sidechain is more exposed than in the deoxy tetramer. With respect to the protein interactions, the mutation stabilizes the oxy tetramer and destabilizes the deoxy tetramer. There are both inter and intrasubunit contributions. As to the intersubunit terms, Tyr $\alpha 42$ (C7) does indeed stabilize the deoxy form in accord with the analysis of Morimoto et al. (1971). By contrast, Asn $\alpha 97$ (G4) favors the oxy form by a relatively small amount. The interaction with Asp $\alpha 94$ (G1) is unfavorable in both the deoxy and oxy form ; i.e., the free energy of interaction between Asp $\alpha 94$ and Asp $\beta 99$ is destabilizing in both tetramers, so that the

Table 1

Free Energy for the Mutation Asp G1(99) $\beta \rightarrow$ Ala^a

Contribution	$\Delta G(\text{deoxy})$	$\Delta G(\text{oxy})$	$\Delta\Delta G(\text{oxy-deoxy})$
Solvent	46.0	68.5	22.5
Protein ^b	20.0	-8.0	-28.0
Asp G1(99) β_2 ²	-8.8	-11.0	-2.2
Inter (α_1)	<u>2.8</u>	<u>-24.4</u>	<u>-27.2</u>
Tyr C7(42)	8.4	-4.3	-12.7
Asp G1(94)	-22.0	-44.4	-22.4
Val G3(96)	1.6	7.1	5.5
Asn G4(97)	9.7	13.0	3.3
Intra (β_2)	<u>26.1</u>	<u>27.4</u>	<u>1.3</u>
His FG4(97)	-2.1	-3.3	-1.2
Pro G2(100)	8.2	5.4	-2.8
Glu G3(101)	-11.2	6.9	18.1
Asn GA(102)	14.3	10.1	-4.2
TOTAL	<u>66.0</u>	<u>60.5</u>	<u>-5.5</u>

- (a) All value in kcal/mole are given for one $\alpha_1\beta_2$ interface ; a term in ΔG with a positive sign corresponds to the fact that the given contribution destabilizes the mutant (Ala) relative to the wild type (Asp). When the effect of the Asp residue by itself is discussed in the text, stabilizing contributions have a positive sign.
- (b) Only the residues which contribute more than 1.5 kcal/mole to both the deoxy and oxyforms are listed.
- (c) Internal energy contribution.

replacement by the nonpolar Ala stabilizes the deoxy tetramer. Also of interest are the contributions that arise from within the β_2 subunit, which are by definition the result of tertiary structural changes that accompany the quaternary transition. All the residues involved are close to the mutated residue Asp α_99 and the largest contribution involves Glu β_{101} . Apparently, the Asp β_{99} / Glu β_{101} interaction is stabilizing in the oxy tetramer and destabilizing in the deoxy tetramer.

It is evident that the free energy simulations provide new insights into the nature of the interactions in proteins and the possible consequences of mutations. Although free energy simulations are a recent development in molecular dynamics so that their reliability is not fully established, it is likely that even if the quantitative values obtained here are not correct, the qualitative insights are still of interest. It is clear that a relatively small overall change in free energy may involve contributions from several large terms. Also, the balance between protein-protein and protein-solvent interactions plays an essential role. Finally, the intrasubunit contribution to cooperativity has not been considered previously.

CONCLUSION

Molecular dynamics is now playing an important role in the study of the properties of macromolecules of biological interest. It is also being used effectively in the analysis of experimental data and, in particular, has been shown to provide a new approach to structure determination by NMR and x-ray crystallography. Because molecular dynamics simulations are relatively new they have so far been employed primarily by theoreticians. It is to be hoped that experimentalists, as well, will begin to use molecular dynamics as a research tool for obtaining a deeper understanding of the biomolecules with which they work.

ACKNOWLEDGEMENTS

I wish to thank my collaborators who have contributed to the specific studies described here. They include C.L. Brooks III, A.T. Brünger, C.M. Clore, C.M. Dobson, R. Elber, J. Gao, A.M. Gronenborn, T. Ichiye, K. Kuczera, J. Kuriyan, R.M. Levy, E.T. Olejniczak, G.A. Petsko and B. Tidor. The work was supported in part by grants from the National Science Foundation and the National Institutes of Health. The present text is essentially that published previously in other reviews.

REFERENCES

- Agmon, M. and Hopfield, J.J., 1983, CO Binding to Heme Proteins : a Model for Barrier Height Distributions and Slow Conformational Changes, *J. Chem. Phys.*, 79:2042.
- Ansari, A., Berendzen, J., Bowne, S.F. Frauenfelder, H., Iben, I.E.T., Sauke, T.B., Shyamsunder, E. and Young, R.D., 1985, Protein States and Proteinquakes, *Proc. Natl. Acad. Sci. USA* 82:5000.
- Artymiuk, P.J., Blake, C.C.F., Grace, D.E.P., Oatley, S.J., Phillips, D.C. and Sternberg, M.J.E., 1979, Crystallographic Studies of the Dynamic Properties of Lysozyme, *Nature* 280:563.

- Austin, R.H., Beeson, K.W., Eisenstein, L., Frauenfelder, H. and Gunsalus, I.C., 1975, Dynamics of Ligand Binding to Myoglobin, *Biochemistry* 14:5355.
- Basford, D., Weaver, D.L. and Karplus, M., 1984, Diffusion-Collision Model for the Folding Kinetics of the λ -Repressor Operator-Binding Domain, *J. Biomol. Struct. Dyn.* 1:1243.
- Bialek, W. and Goldstein, R.F., 1985, Do Vibrational Spectroscopies Uniquely Describe Protein Dynamics ? The Case for Myoglobin, *Biophys. J.*, 8:1027.
- Blake, C.C.F., Pulford, W.C.A. and Artymiuk, P.J., 1983, X-ray Studies of Water in Crystals of Lysozyme, *J. Mol. Biol.* 167:693.
- Brooks, B.R. and Karplus, M., 1983, Harmonic Dynamics of Proteins : Normal Modes and Fluctuations in Bovine Pancreatic Trypsin Inhibitor, *Proc. Natl. Acad. Sci. USA* 80:6571.
- Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S. and Karplus, M., 1983, CHARMM : A Program for Macromolecular Energy Minimization, and Dynamics Calculations, *J. Comp. Chem.* 4:187.
- Brooks III, C.L., Brünger, A.T. and Karplus, M., 1985, Active Site Dynamics in Protein Molecules : A Stochastic Boundary Molecular-Dynamics Approach, *Biopolymers* 24:843.
- Brooks III, C.L., Karplus, M. and Pettitt, B.M., 1987, *Proteins : A Theoretical Perspective of Dynamics, Structure and Thermodynamics*, Adv. Chem. Phys. LXXI, John Wiley & Sons, New York.
- Brünger, A.T., Brooks III, C.L. and Karplus, M., 1984, Stochastic Boundary Conditions for Molecular Dynamics Simulations of ST2 Water, *Chem. Phys. Letters* 105:495.
- Brünger, A.T., Brooks III, C.L. and Karplus, M., 1985, Active Site Dynamics of Ribonuclease, *Proc. Natl. Acad. Sci. USA* 82:8458.
- Brünger, A.T., Clore, G.M., Gronenborn, A.M. and Karplus, M., 1986, Three-Dimensional Structure of Proteins Determined by Molecular Dynamics with Interproton Distance Restraints : Application to Crambin, *Proc. Natl. Acad. Sci. USA* 83:3801.
- Brünger, A.T., Kuriyan, J. and Karplus, M., 1987a, Crystallographic R Factor Refinement by Molecular Dynamics, *Science* 235:458.
- Brünger, A.T., Campbell, R.L., Clore, G.M., Gronenborn, A.M., Karplus, M., Petsko, G.A. and Teeter, M.M., 1987b, Solution of a Protein Crystal Structure with a Model Obtained from NMR Interproton Distance Restraints, *Science* 235:1049.
- Brünger, A.T., 1989, A Memory-Efficient Fast Fourier Transformation Algorithm for Crystallographic Refinement on Supercomputers, *Acta Cryst.* A45:42.
- Brünger, A.T., Karplus, M. and Petsko, G.A., 1989, Crystallographic Refinement by Simulated Annealing : Application to Crambin, *Acta Cryst.* BA45:50.
- Bunn, H.F. and Forget, B.G., 1980, "Hemoglobin : Molecular, Genetic Clinical Aspects", Saunders, New York.

- Campbell, I.D., Dobson, C.M., Moore, G.R., Perkins, S.J. and Williams, R.J.P., 1976, Temperature Dependent Molecular Motion of a Tyrosine Residue of Ferrocycochrome C, FEBS Lett. 70:96.
- Campbell, I.D., Dobson, C.M. and Williams, R.J.P., 1978, Structures and Energetics of Proteins and their Active Sites, Adv. Chem. Phys. 39:55
- Chandrasekhar, S., 1943, Stochastic Problems in Physics and Astronomy, Rev. Mod. Phys., 15:1.
- Chothia, C. and Lesk, A.M., 1985, Helix Movements in Proteins, TIBS 10:116.
- Clare, G.M., Gronenborn, A.M., Brünger, A.T. and Karplus, M., 1985, Solution Conformation of a Heptadecapeptide Comprising the DNA Binding Helix F of the Cyclic AMP Receptor Protein of *Escherichia coli*: Combined Use of ^1H Nuclear Magnetic Resonance and Restrained Molecular Dynamics, J. Mol. Biol. 186:435.
- Clare, G.M. and Gronenborn, A.M., 1989, Determination of Three-Dimensional Structures of Proteins and Nucleic Acids in Solution by Nuclear Magnetic Resonance Spectroscopy, CRC Crit. Rev. Biochem., in press.
- Debrunner, P.G. and Frauenfelder, H., 1982, Dynamics of Proteins, Ann. Rev. Phys. Chem. 33:283.
- Desmeules, P.J. and Allen, L.C., 1980, Strong, Positive-Ion Hydrogen Bonds : The Binary Complexes Formed from NH_3 , OH_2 , FH , PH_3 , SH_2 and ClH , J. Chem. Phys. 72:4731.
- Dickerson, R.E. and Geis, I., 1983, "Hemoglobin : Structure, Function, Evolution, and Pathology", Benjamin/Cummings, Menlo Park.
- Dobson, C.M. and Karplus, M., 1986, Internal Motion of Proteins : Nuclear Magnetic Resonance Measurements and Dynamic Simulations, in "Methods in Enzymology", 131, C.H.W. Hirs and S.N. Timasheff, eds., Academic Press, Inc., New York.
- Doster, W., Cusack, S. and Petry, W., 1989, Dynamical Transition of Myoglobin Revealed by Inelastic Neutron Scattering, Nature 337:734.
- Elber, R. and Karplus, M., 1987, Multiple Conformational States of Proteins : A Molecular Dynamics Analysis of Myoglobin, Science 235:318.
- Fermi, G. and Perutz, M.F., 1981, "Haemoglobin and Myoglobin", Atlas of Molecular Structures in Biology : 2, Clarendon, Oxford.
- Fersht, A.R., 1987, The Hydrogen Bond in Molecular Recognition, Trends Biochem. Sci. 12:301.
- Fleischman, S.H., Tidor, B., Brooks III, C.L. and Karplus, M., Free Energy Simulation Methodology, J. Comp. Chem. to be published.
- Frauenfelder, H., Petsko, G.A. and Tsernoglou, D., 1979, Temperature-Dependent X-Ray Diffraction as a Probe of Protein Structural Dynamics, Nature 280:558.
- Gao, J. Kuczera, K., Tidor, B. and Karplus, M., 1989, Thermodynamic Analysis of a Cooperativity Mutant in Hemoglobin, Science, to be published.

- Gelin, B.R., Lee, A.W.-M. and Karplus, M., 1983, Hemoglobin Tertiary Structural Change on Ligand Binding : Its Role in the Cooperative Mechanism, *J. Mol. Biol.* 171:489.
- Glover, I., Haneef, I., Pitts, J., Wood, S., Moss, D., Tickle, I. and Blundell, T., 1983, Conformational Flexibility in a Small Globular Harmane : X-Ray Analysis of Avian Pancreatic Polypeptide at 0.98-Å Resolution, *Biopolymers* 22:293.
- Gurd, F.R.N. and Rothgeb, J.M., 1979, Motions in Proteins, *Adv. Prot. Chem.* 33:73.
- Hansen, J.P. and McDonald, I.R., 1976, "Theory of Simple Liquids", Academic Press, New York.
- Hartmann, H., Parak, F., Steigemann, W., Petsko, G.A., Ringe Ponzi, D. and Frauenfelder, H., 1982, Conformational Substates in a Protein : Structure and Dynamics of Metmyoglobin at 80°K, *Proc. Natl. Acad. Sci. USA* 79:4967.
- Hendrickson, W.A. and Teeter, M.M., 1981, Structure of the Hydrophobic Protein Crambin Determined Directly from the Anomalous Scattering of Sulphur, *Nature* 290:107.
- Honig, B., Hudson, B., Sykes, B.D. and Karplus, M., 1971, Ring Orientation in β -Ionone and Retinals, *Proc. Natl. Acad. Sci. USA* 68:1289.
- Ichiye, T., Olafson, B.D., Swaminathan, S. and Karplus, M., 1986, Structure and Internal Mobility of Proteins : A Molecular Dynamics Study of Hen Egg White Lysozyme, *Biopolymers* 25:1909.
- Jack, A. and Levitt, M., 1978, Refinement of Large Structures by Simultaneous Minimization of Energy and R Factor, *Acta Cryst.* A34:931.
- James, M.N.G. and Sielecki, A.R., 1983, Structure and Refinement of Penicillopepsin at 1.8 Å Resolution, *J. Mol. Biol.* 163:299.
- Karplus, M. and Kushick, J.N., 1981, Method for Estimating the Configurational Entropy of Macromolecules, *Macromolecules* 14:325.
- Karplus, M. and McCammon, J.A., 1981, The Internal Dynamics of Globular Proteins, *CRC Crit. Rev. Biochem.* 9:293.
- Karplus, M. and McCammon, J.A., 1983, Dynamics of Proteins : Elements and Function, *Ann. Rev. Biochem.* 53:263.
- Kearle, P., 1977, Ion Thermochemistry and Solvation from Gas Phase Ion Equilibria, *Ann. Rev. Phys. Chem.* 28:445.
- Kirkpatrick, S., Gelatt, Jr., C.D. and Vecchi, M.P., 1983, Optimization by Simulated Annealing, *Science* 220:671.
- Kirkwood, J.G., 1935, Statistical Mechanics of Fluid Mixtures, *J. Chem. Phys.* 3:300.
- Kirkwood, J.G. and Boggs, E.M., 1942, The Radial Distribution Function in Liquids, *J. Chem. Phys.* 10:394.
- Konnert, J.H. and Hendrickson, W.A., 1980, A Restrained-Parameter Thermal-Factor Refinement Procedure, *Acta Cryst.* A36:344.

- Kuriyan, J., Petsko, G.A., Levy, R.M. and Karplus, M., 1986, Effect of Anisotropy and Anharmonicity on Protein Crystallographic Refinement : An evaluation by Molecular Dynamics, *J. Mol. Biol.* 190:227.
- Kuriyan, J., Karplus, M. and Petsko, G.A., 1987, Estimation of Uncertainties in X-Ray Refinement Results by Use of Perturbed Structures, *Proteins* 2:1.
- Lesk, A.M. and Chothia, C., 1980, How Different Amino Acid Sequences Determine Similar Protein Structures : The Structure and Evolutionary Dynamics of the Globins, *J. Mol. Biol.* 136:225.
- Levitt, M., Sander, C. and Stern, P.S., 1985, Protein Normal-Mode Dynamics : Trypsin Inhibitor, Crambin, Ribonuclease and Lysozyme, *J. Mol. Biol.* 181:423.
- Levy, R.M. and Karplus, M., 1979, Vibrational Approach to the Dynamics of an α -Helix, *Biopolymers* 18:2465.
- Levy, R.M., Perahia, D. and Karplus, M., 1982, Molecular Dynamics of an α -Helical Polypeptide : Temperature Dependence and Deviation from Harmonic Behavior, *Proc. Natl. Acad. Sci. USA* 79:1346.
- Levy, R.M., Sheridan, R.P., Keepers, J.W., Dubey, G.S., Swaminathan, S. and Karplus, M., 1985, Molecular Dynamics of Myoglobin at 298°K : Results from a 300 ps Computer Simulation, *Biophys. J.* 48:509.
- Levy, R.M., Karplus, M. and Wolynes, P.G., 1981, NMR Relaxation Parameters in Molecules with Internal Motion : Exact Langevin Trajectory Results Compared with Simplified Relaxation Models, *J. Am. Chem. Soc.* 103:5998.
- Marfey, P.S., Uziel, M. and Little, J., 1965, Reaction of Bovine Pancreatic Ribonuclease A with 1,5-Difluoro-2,4-dinitrobenzene, *J. Biol. Chem.* 240:3270.
- Marquart, M., Deisenhofer, D., Huber, R. and Palm, W., 1980, Crystallographic Refinement and Atomic Models of the Intact Immunoglobulin Molecule Kol and its Antigen-Binding Fragments at 3.04 Å and 1.9 Å Resolution, *J. Mol. Biol.* 141:369.
- Matthew, J.B. and Richards, F.M., 1982, Anion Binding and pH-Dependent Electrostatic Effects in Ribonuclease, *Biochemistry* 21:4989.
- McCammon, J.A., Gelin, B.R. and Karplus, M., 1977, Dynamics of Folded Proteins, *Nature* 267:585.
- McCammon, J.A., Wolynes, P.G. and Karplus, M., 1979, Picosecond Dynamics of Tyrosine Side Chains in proteins, *Biochemistry* 18:927.
- Morimoto, H., Lehmann, H. and Perutz, M.F., 1971, Molecular Pathology of Human Haemoglobin : Stereochemical Interpretation of Abnormal Oxygen Affinities, *Nature* 232:408.
- McQuarrie, D.A., 1976, "Statistical Mechanics", Harper & Row, New York.
- Moss, D.S. and Morffew, A.J., 1982, Restrain : A Restrained Least Squares Refinement Program for Use in Protein Crystallography, *Comput. & Chem.* 6:1.

- Navia, M.A., Fitzgerald, P.M.D., McKeever, B.M., Leu, C.-T., Heimbach, J.C., Herber, W.K., Sigal, I.S., Darke, P.L. and Springer, J.P., 1989, Three-Dimensional Structure of Aspartyl Protease from Human Immunodeficiency Virus HIV-1, *Nature* 337:615.
- Noggle, J.H. and Schirmer, R.E., 1981, "The Nuclear Overhauser Effect", Academic Press, New York.
- Northrup, S.H., Pear, M.R., Lee, C.-Y., McCammon, J.A. and Karplus, M., 1982, Dynamical Theory of Activated Processes in Globular Proteins, *Proc. Natl. Acad. Sci. USA* 79:4035.
- Olejniczak, E.T., Dobson, C.M., Levy, R.M. and Karplus, M., 1984, Motional Averaging of Proton Nuclear Overhauser Effects in Proteins. Predictions from a Molecular Dynamics Simulation of Lysozyme, *J. Am. Chem. Soc.* 106:1923.
- Olejniczak, E.T., Poulsen, F.M. and Dobson, C.M., 1981, Proton Nuclear Overhauser Effects and Protein Dynamics, *J. Am. Chem. Soc.* 103:6574.
- Parak, F., Knapp, E.W. and Kucheida, D., 1982, Protein Dynamics. Mössbauer Spectroscopy on Deoxymyoglobin Crystals, *J. Mol. Biol.* 161:177.
- Pauling, L., Corey, R.B. and Branson, H.R., 1951, The Structure of Proteins : Two Hydrogen-Bonded Helical Configurations of the Polypeptide Chain, *Proc. Natl. Acad. Sci. USA* 37:205.
- Perutz, M.F., 1970, Haem-Haem Interaction and the Problem of Allostery, *Nature* 228:726.
- Petsko, G.A. and Ringe, D., 1984, Fluctuations in Protein Structure from X-ray Diffraction, *Ann. Rev. Biophys. Bioeng.* 13:331.
- Pettigrew, D.W., Romeo, P.H., Tsapis, A., Thillet, J., Smith, M.L., Turner, B.W. and Ackers, G.K., 1982, Probing the Energetics of Proteins through Structural Perturbation : Sites of Regulatory Energy in Human Hemoglobin, *Proc. Natl. Acad. Sci. USA* 79:1849.
- Phillips, D.C., 1981, Closing Remarks, in "Biomolecular Stereodynamics", R.H. Sarma, ed., Adenine, New York.
- Poulsen, F.M., Hoch, J.C. and Dobson, C.M., 1980, A Structural Study of the Hydrophobic Box Region of Lysozyme in Solution Using Nuclear Overhauser Effects, *Biochemistry* 19:2597.
- Ramachandran, G.N., Ramakrishnan, C. and Sasisekharan, V., 1963, Stereochemistry of Polypeptide Chain Configurations, *J. Mol. Biol.* 7:95.
- Shih, H.H.-L., Brady, J. and Karplus, M., 1985, *Proc. Natl. Acad. Sci. USA* 82:1697.
- Smith, J.L., Hendrickson, W.A., Honzatko, R.B. and Sheriff, S., 1986, Structural Heterogeneity in Protein Crystals, *Biochemistry* 25:5018.
- Solomon, I., 1955, Relaxation Processes in a System of Two-Spins, *Phys. Rev.* 99:559.
- Stein, D.L., 1985, A model of Protein Conformational Substates, *Proc. Natl. Acad. Sci. USA* 82:3670.

- Stillinger, F.H. and Rahman, A., 1974, Improved Simulation of Liquid Water by Molecular Dynamics, *J. Chem. Phys.* 60:1545.
- Stillinger, F.H. and Weber, T.A., 1982, Hidden Structure in Liquids, *Phys. Rev. A* 25:978.
- Stillinger, F.H. and Weber, T.A., 1984, Packing Structures and Transitions in Liquids and Solids, *Science* 225:983.
- Sussman, J.L., Holbrook, S.R., Church, G.M., Kim, S.-H., 1977, A Structure-Factor Least-Squares Refinement Procedure for Macromolecular Structures Using Constrained and Restrained Parameters, *Acta Cryst.* A33:800.
- Swaminathan, S., Ichiye, T., van Gunsteren, W.F. and Karplus, M., 1982, Time Dependence of Atomic Fluctuations in Proteins : Analysis of Local and Collective Motions in Bovine Pancreatic Trypsin Inhibitor, *Biochemistry* 21:5230.
- Toulouse, G., 1984, Progrès récents dans la physique des systèmes désordonnés, *Helv. Phys. Acta* 57:459.
- van Gunsteren, W.F. and Berendsen, M.J.C., 1977, Algorithms for Macromolecular Dynamics and Constraint Dynamic, *Mol. Phys.* 34:1311.
- van Gunsteren, W.F. and Karplus, M., 1982, Effect of Constraints on the Dynamics of Macromolecules, *Macromolecules* 15:1528.
- Verlet, L., 1967, Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules, *Phys. Rev.* 159:98.
- Wagner, G. and Wüthrich, K., 1982, Amide Proton Exchange and Surface Conformation of the Basic Pancreatic Trypsin Inhibitor in Solution, *J. Mol. Biol.* 160:343.
- Weatherall, D.J., Clegg, J.B., Callender, S.T., Wells, R.M.G., Gale, R.E., Huenhs, E.R., Perutz, M.F., Viggiano, G. and Ho, C., 1977, Haemoglobin Radcliffe ($\alpha 2\beta 2$ 99) (G1,Ala) : A High Oxygen-Affinity Variant Causing Familial Polycythaemia, *British J. Haematology* 35:177.
- Weber, P.C., Salemme, F.R., Lin, S.H., Konishi, Y. and Scheraga, H.A., 1985, Preliminary Crystallographic Data for Cross-Linked (Lysine⁷-Lysine⁴¹) - Ribonuclease A, *J. Mol. Biol.* 181:453.
- Wlodawer, A., 1985, Structure of Bovine Pancreatic Ribonuclease by X-Ray and Neutron Diffraction, in *Biological Macromolecules and Assemblies : Vol. 2, Nucleic Acids and Interactive Proteins*, F.A. Jurnak and A. McPherson, eds., Wiley, New York.
- Wüthrich, K., 1989, The Development of Nuclear Magnetic Resonance Spectroscopy as a Technique for Protein Structure Determination, *Acc. Chem. Res.* 22:36.
- Wyckoff, H.W., Hirs, C.H.W. and Timasheff, S.N., Eds, 1985, "Diffraction Methods for Biological Macromolecules", Part B. *Methods Enzymology*, 115.
- Yu, H.-A., Karplus, M. and Hendrickson, W.A., 1985, Restraints in Temperature- Factor Refinement for Macromolecules : an Evaluation by Molecular Dynamics, *Acta Cryst.* B41:191.

Ziman, J.M., 1979, "Models of Disorder : The Theoretical Physics of Homogeneously Disordered Systems", Cambridge University Press, New York.

DISCUSSION

RULLMANN - You showed how a protein is moving in time through a large number of substates. Is it possible to say anything about the statistical or topological distribution of these substates, and could such knowledge be used to improve our understanding of the behaviour of macromolecular systems ?

KARPLUS - As I indicated in my talk, substates that are very similar are accessed on a subpicosecond timescale. They differ only locally (0.2 Å rms). As the time interval and rms structural difference increase, more extensive parts of the protein are involved. At the long-time limit (on the order of 150 ps) any pair of substates differ from each other by changes that are distributed throughout the protein. As yet it has not been possible to evaluate quantitatively the relative energies of a large enough population of substates to permit a statistical evaluation of their contributions.

DYMEK - The idea of using 60 CO's to analyze the path of CO from the heme pocket to outside the myoglobin by "turning off" the CO repulsions is a good idea. But doesn't the presence of a CO in the myoglobin affect its fluctuations in a way that changes the potentials that the other CO molecules see ?

KARPLUS - In applying the time-dependent Hartree approximation to photodissociation of CO myoglobin, the assumption is made that the effect of a CO molecule on the protein fluctuations is small. An analysis of the protein in the presence of 60 CO molecules suggest that its dynamics are very similar to free myoglobin dynamics. It does not appear to be true that when one CO goes over a barrier it is easier for a second one to follow in a correlated fashion. All of these results support the use of the time-dependent Hartree method for obtaining CO pathways as a good first approximation.

WARSHEL - Basically your study presented the different contribution to the very large (- 70K cal/mol) energy of charging an ionizable group in a protein. Now while this is very important, earlier (1984) you argued in response to a related calculation that no one should calculate such numbers since no one measures these energies.

KARPLUS - It is true that if one can calculate the energy difference between a charged group in solution and the protein directly, this permits one to cancel out the large contribution which comes from solvation of the gas phase ion. Methods based on continuum electrostatics are useful in that regard. However, now that it is possible to make many mutations in proteins (which was not true in 1984 when the discussion to which you refer took place ; I presume you are referring to the Molecular Dynamics Workshop in North Carolina), it is more important to dissect the various contributions so one can determine what can happen when one removes a charged group which can interact strongly with other charged groups in the protein.

GRESH - The existence of a favourable interaction between Glu 101 and the mutatest aspartate is unexpected. Could the dielectric screening be incriminated for it ?

KARPLUS - Two charged groups can have a favorable interaction (in the present calculation this corresponds to a destabilizing contribution when the aspartic acid is changed to alanine) if the orientation of the two groups is appropriate. One must remember that one is not simply dealing with a single negative charge, but a distribution of charges which adds up to minus one. Since dielectric screening comes into the present calculation indirectly (i.e., due to the contribution of solvent interactions), the results that I reported cannot be based on this.

BUCKINGHAM - I have three remarks :

1. Your simulations on the hemoglobin mutants shows that the significant free energy differences emerge as the differences between large numbers. How accurate do you estimate these differences to be ?
2. The hydrogen bond is known to be non-additive, and this is thought to be due to polarization. Do you incorporate non-additive interactions in your computations ?
3. One could test for cooperative effects in the interaction of CO molecules with the protein by doing the simulation with different numbers of CO's. If the rate of reaction per CO is independent of the number of CO's, then the cooperative effect is negligible.

KARPLUS - Your first question concerns the accuracy of the overall free energy change which, as I pointed out, is a difference between large numbers. There is a significant uncertainty in the overall free energy change (perhaps even of the magnitude of the free energy difference itself). However, the qualitative features of the large individual contributions are correctly given by the method. The work which I

described is concerned with demonstrating that there exists these large contributions, and with suggesting that they may appear directly when appropriate mutations are made.

As to the nonadditivity of hydrogen bonds due to polarization, that is not included in the force field we are using. There is, however, some nonadditivity of hydrogen bond interactions with water in that the orientation effects on the water lead to a larger favorable contribution to the free energy than from the second hydrogen bond than the first.

Your idea for testing possible cooperative effects in the CO simulation by using different numbers of CO's is interesting. It certainly would be one way of approaching the problem.

DURUP - Back to the calculation with the 60 carbon monoxide molecules : I think the model does not take into account the possibility that the CO molecule would force its way through the protein just like you do in a tight crowd with your shoulders. Would it not be possible to allow, for each time span of a few picoseconds, the protein to feel only the interaction with one CO chosen by some criterion (largest interaction energy at that given moment), and thereafter to give its chance to another CO molecule, etc.?

KARPLUS - It is indeed true, as I pointed out in my talk, that cooperative effects are neglected. From examining the simulations, it appears that the CO does not "force" its way through the protein, but that most commonly the protein fluctuations permit the CO to get through a region which normally has a high barrier. There are certainly alternative approaches which might refine the methodology we employed in this pioneering study. Your suggestion describes one of them and it would be interesting to implement.

WIPFF - What would be the effect of the surrounding of protein on these dynamics ? How for instance do the dynamics of globin compare in vacuo, in the crystalline state, and in aqueous solution ?

KARPLUS - The question of the effect of the surroundings of a protein on its dynamics is one that is often asked. From experiments such as incoherent neutron scattering and from comparisons of simulations in different environments, it appears that the interior motions of a protein are little affected by the environment. The surface motions are significantly affected by the environment. The surface motions are significantly affected

either by interactions with neighboring molecules in the crystal or with the solvent. From simulations at varying temperatures it appears that the temperature plays an important direct role in altering the dynamics, though the environment effects of the temperature (e.g., freezing of the solvent) may also contribute. We have not made a detailed study of myoglobin and dynamics in different environments.

ON TESTING THEORETICAL MODELS BY COMPARISON OF CALCULATED WITH
EXPERIMENTAL DATA

W.F. VAN GUNSTEREN

*Department of Physical Chemistry, University of Groningen, Nijenborgh 16,
9747 AG GRONINGEN, The Netherlands, and
Department of Physics, Free University, P.O. Box 7161, 1007 MC
AMSTERDAM, The Netherlands.*

Abstract

Good agreement between calculated and experimental data does not necessarily mean that the underlying theoretical model is correct. The good agreement may be due to compensation of errors. A number of examples of such a situation are discussed.

1. Introduction

In the field of molecular modelling and computational chemistry theoretical models of molecular systems are generally based on the basic physical laws and interactions. Molecular models that are used to describe complex molecular systems, like solvated proteins, often contain many parameters and a large number of degrees of freedom, e.g. the positions of the atoms in the system. The assumptions implicit in a particular model and the parameter values that are used, are generally validated by a comparison of theoretically predicted with experimentally measured properties of the system under consideration. The results of such a comparison between theory and experiment can be classified as follows.

A. *Agreement between theory (model) and experiment*

The good agreement may be due to one or more of the following reasons:

1. The *model is correct*, that is, any other assumption used to derive the model, or any other choice of parameter values would give bad agreement with experiment.
2. The *property* that is compared is *insensitive to the assumptions* or *parameter values* of the model, that is, whatever parameter values

are used in the model calculation the agreement with experiment will be good.

3. *Compensation of errors* occurs, either by chance or by fitting of the model parameters to the desired properties.

B. *No agreement between theory (model) and experiment*

This may be due to the following reasons:

1. The *model* is *not correct*, the experiment is.
2. The model is correct, the *experimental data* are *not correct*.

In this paper we will give a number of examples of case A3. We show that when only a few numbers or properties of highly dimensional systems or models containing many parameters are compared to experimental data, it may easily occur that good agreement is obtained due to compensation of errors. In these cases the good agreement between calculated and predicted properties does not imply that the theoretical model is correct.

2. Calculation of the crystallographic R-factor

The crystallographic *reliability factor* or R-value is defined as

$$R = \frac{\sum_{hkl} W(hkl) ||F_{obs}(hkl)| - k_{sc}|F_{calc}(hkl)||}{\sum_{hkl} |F_{obs}(hkl)|} * 100 \% \quad (2.1)$$

where the calculated (F_{calc}) or observed (F_{obs}) structure factor is defined as the Fourier transform over the unit cell

$$F(hkl) = \text{constant} * \iiint \rho(xyz) e^{2\pi i(hx+ky+lz)} dx dy dz \quad (2.2)$$

of the electron density $\rho(xyz)$. In an X-ray diffraction experiment reflection intensities

$$I(hkl) \sim |F(hkl)|^2 \quad (2.3)$$

are measured. Weight factors in reciprocal space are denoted by $W(hkl)$.

From definition (2.1) it is clear that the better the agreement between

$|F_{calc}|$ and $|F_{obs}|$, the lower the value of R . However, the R -value also depends critically on the number of observations and parameters in (2.1), that is, the R -value depends on

- the resolution (d) range of the diffraction pattern
- the number N_{refl} of observed reflections
- the number N_{sc} of scaling factors k_{sc}
- the parameters in the molecular model that is used to compute F_{calc} , etc.

In table I it is illustrated how the R -value may be reduced by reducing the number of observations (resolution range d , number of reflections N_r) or by increasing the number of adjustable parameters (number of scaling factors N_{sc}). For one protein structure the R -value may attain values between 35 % and 24 % depending on the choice of the parameters involved in (2.1). This illustrates the - of course known - fact that a low R -value, that is, good agreement with experimental data according to (2.1), does not necessarily imply that the correct molecular structure has been found.

Table I
Dependence of R -value upon number of parameters^{a)}

resolution range d (Å)	number of reflections N_{refl}	number of scaling factors N_{sc}	R -value (%) ^{a)}
2.5-10	all	1	35.0
2.5-7	all	1	31.5
2.5-7	all	20	28.7
2.5-7	$I > 2\sigma$	20	24.0

^{a)}Values for the X-ray structure of α -bungarotoxin (Love & Stroud, 1986; Brookhaven Protein Data Bank, Bernstein et al., 1977; J. Finan private communication).

3. Density of a cytidine derivative crystal

In table II the dimensions are shown of the unit cell of a cytidine derivative (Verdegaal et al., 1981) for which the crystal structure has been determined at 1 atm pressure and two different temperatures, viz. 113 K

(Verdegaal et al., 1981) and 289 K (James & Sielecki, unpublished work). Molecular dynamics (MD) simulations of 4 unit cells using crystalline periodic boundary conditions and the constant pressure, constant temperature, variable volume algorithm of Berendsen et al. (1984) have been performed at the two mentioned temperatures in order to compare model and experimental properties (Van Gunsteren & Berendsen, 1985).

Table II
Densitv of cytidine derivative crvstals^{a)}

4 unit cells dimensions (nm) or volume (nm ³)	<u>temperature: 113K</u>		<u>temperature: 289K</u>	
	experiment	MD	experiment	MD
2a	1.788	1.792	1.778	1.790
2b	2.065	2.050	2.077	2.064
c	2.758	2.772	2.864	2.822
volume (4abc)	10.18	10.18	10.56	10.43

^{a)}(Van Gunsteren & Berendsen, 1985; Verdegaal et al. 1981)

The unit cell volume is very well reproduced; a deviation of 0.0 % (113 K) and 1 % (289 K). However, this nearly perfect agreement is due to compensation of errors: at 113 K the model yields too small a value for *b* and too large values for *a* and *c*; at 289 K *a* is too large and *b* and *c* are too small. By comparing three quantities in stead of one quantity, the agreement with experiment is reduced.

4. Conservation of total energy when integrating Newton's equations of motion

In this section we give an example of compensation of errors leading to incorrect conclusions with respect to the efficiency of two different integration algorithms which are used in molecular dynamics (MD) simulations.

In Newtonian dynamics the total energy (E_{tot}), the sum of the kinetic (E_{kin}) and the potential (E_{pot}) energy, is conserved:

$$E_{tot} = \frac{1}{2} \sum_{i=1}^N m_i \vec{v}_i^2 + V(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N) = \text{constant} \quad (4.1)$$

This means that the better the integration of Newton's equations is performed, the smaller the fluctuation in the total energy,

$$\Delta E_{tot} = \langle [E_{tot} - \langle E_{tot} \rangle]^2 \rangle^{1/2} \quad (4.2)$$

will be. Here averaging over the trajectory is denoted by $\langle \dots \rangle$. Therefore, one may think that the size of ΔE_{tot} may be used to judge the performance of an integration algorithm.

Table III
Conservation of total energy in MD simulations^{a)}

		<u>Fluctuation ΔE_{tot} of the total energy (kcal/mol)</u>		
Algorithm	bond vibrations	$\Delta t = 0.5$ fs	$\Delta t = 1$ fs	$\Delta t = 2$ fs
Beeman	yes	0.004	0.30	1.1
Verlet	yes	0.3	1.2	4.3
Verlet/SHAKE	no	0.09	0.2	1.0

^{a)}Data for MD simulations of the protein BPTI, taken from Levitt (1983).

In table III the fluctuation of the total energy is given for two different algorithms, the Beeman and the Verlet algorithm. The data are from Levitt (1983). Since the Beeman algorithm produces the smallest ΔE_{tot} , Levitt concludes that it is better than the Verlet algorithm. However, it can be shown (Aqvist et al., 1985; Berendsen & Van Gunsteren, 1986) that the Beeman algorithm (Beeman, 1976)

$$x(t+\Delta t) = x(t) + v(t)\Delta t + [4a(t) - a(t-\Delta t)](\Delta t)^2/6 \quad (4.3.a)$$

$$v(t+\Delta t) = v(t) + [2a(t+\Delta t) + 5a(t) - a(t-\Delta t)]\Delta t/6 \quad (4.3.b)$$

produces exactly the same trajectory as the Verlet algorithm (Verlet, 1967)

$$x(t+\Delta t) = 2x(t) - x(t-\Delta t) + a(t)(\Delta t)^2 \quad (4.4.a)$$

$$v(t) = [x(t+\Delta t) - x(t-\Delta t)]/(2\Delta t). \quad (4.4.b)$$

Positions $x(t)$, velocities $v(t)$ and acceleration $a(t)$ are functions of time t over which is integrated in steps of size Δt . If the Beeman and Verlet trajectories are identical, how can the different ΔE_{tot} values in the columns of Table III be explained?

The identical trajectories yield identical ΔE_{pot} values, but the Verlet velocity formula (4.4.b) is less accurate than the Beeman formula (4.3.b). This leads to

$$\Delta E_{kin}(\text{Verlet}) > \Delta E_{kin}(\text{Beeman}) \quad (4.5)$$

and because of

$$\Delta E_{pot}(\text{Verlet}) = \Delta E_{pot}(\text{Beeman}) \quad (4.6)$$

to

$$\Delta E_{tot}(\text{Verlet}) > \Delta E_{tot}(\text{Beeman}). \quad (4.7)$$

We note that the Verlet trajectory is independent of formula (4.4.b), since $x(t+\Delta t)$ only depends on previous positions and acceleration, not on the velocity. Relation (4.7) is reflected in the first two lines of table III.

The third line shows the effect on ΔE_{tot} of removing the highest-frequency (bond-length) vibrations from the molecule, which can be done by applying the procedure SHAKE (Ryckaert et al., 1977). Removal of high-frequency motions yield better energy conservation, as expected.

So, the last column $\Delta t = 2$ fs of table III contains an example of obtaining almost identical numbers for different reasons. Beeman with bond vibrations inaccurately integrated ($\Delta t = 2$ fs is too large) yields $\Delta E_{tot} = 1.1$ kcal/mol. When using Verlet, the poor velocity formula (4.4.b) raises ΔE_{tot} to 4.3 kcal/mol, but the trajectory is not changed due to the equivalence of (4.3) and (4.4.a). Upon removal of the bond vibrations the equations of motion are more accurately integrated, which leads to a reduction of ΔE_{tot} to 1.0 kcal/mol. From the first two columns of table III it is observed that the removal of bond vibrations cannot completely cancel the large kinetic energy

fluctuations due to (4.4.b) when Δt is taken smaller. A proper analysis of Levitt's data (1983) leads to the conclusion that the Verlet/SHAKE algorithm yields more accurately integrated trajectories than the Beeman algorithm, whereas a loose comparison of ΔE_{tot} values leads to the opposite (not correct) conclusion.

5. The computation of free enthalpy: adequate sampling

The relative free energy or enthalpy difference $\Delta G(A \rightarrow B) \equiv \Delta G_{BA} \equiv G_B - G_A$ between two states A and B of a molecular system can be obtained by the technique of thermodynamic integration (see e.g. Van Gunsteren & Weiner, 1989). The potential energy $V(\vec{r})$ is made a function of a coupling parameter λ in such a manner that $V(\vec{r}, \lambda_A)$ corresponds to the system in state A, and $V(\vec{r}, \lambda_B)$ to state B. During a simulation the coupling parameter is slowly changed from λ_A to λ_B . If the change is reversibly performed ΔG is obtained as an integral over the configuration space accessible to the system. The reliability of the obtained ΔG value strongly depends on the extent of the sampling of configuration space. Partial information on the adequacy of the sampling can be obtained by changing λ forward from λ_A to λ_B and backward to λ_A and computing the hysteresis along the closed loop

$$\Delta G_{hys} \equiv \Delta G(A \rightarrow B) + \Delta G(B \rightarrow A). \quad (5.1)$$

Table IV
Hysteresis in free enthalpy calculation

Condition ^{a)}	ΔG_{hys}	Obtained ΔG value
$\tau_{MD} \ll \tau_{system}$	0	incorrect
$\tau_{MD} \approx \tau_{system}$	$\neq 0$	incorrect
$\tau_{MD} \gg \tau_{system}$	0	correct

^{a)}The relaxation time of the system (environment) is denoted by τ_{system} . The time period of the MD simulation over which the change from state A to state B is performed is denoted by τ_{MD} .

Three cases can be distinguished (table IV). If the length τ_{MD} of the MD simulation is much longer than the slowest relaxation time τ_{system} of the system, sampling is likely to be sufficient, $\Delta G_{hys} \approx 0$ and the obtained ΔG value is reliable. When $\tau_{MD} \approx \tau_{system}$, sampling is insufficient, the change A-B-A is irreversible, leading to $\Delta G_{hys} \neq 0$ indicating an incorrect ΔG value. If $\tau_{MD} \ll \tau_{system}$, the hysteresis will be nearly zero, since the system cannot adapt itself at all to the change A-B-A. An incorrect ΔG value is obtained despite the observation $\Delta G_{hys} \approx 0$.

The last two cases are illustrated in table V for the process of changing a Ne atom into a Na^+ ion in aqueous solution. For the shorter τ_{MD} the hysteresis is considerable, it is only reduced when $\tau_{MD} > \tau_{water}$ (dielectric relaxation time ≈ 8 ps; rotational correlation time ≈ 2 ps). Nevertheless the averaged free enthalpy estimate ΔG_{av} is remarkably independent of τ_{MD} , due to compensation of errors in the forward and backward integration.

Table V
Hysteresis as a function of integration period^{a)}

length of MD simulation τ_{MD} (ps)	$\Delta G_{hys} = \Delta G(A \rightarrow B) + \Delta G(B \rightarrow A)$ (kJ/mol)	$\Delta G_{av} = \frac{\Delta G(A \rightarrow B) - \Delta G(B \rightarrow A)}{2}$ (kJ/mol)
5	39	-422
10	29	-419
20	18	-420
40	1	-421
80	1	-421

^{a)}Data for the process of changing neon (state A) in aqueous solution (216 H_2O , $\Delta t = 1fs$, $R_c = 0.9nm$) to sodium (state B), taken from Straatsma (1987).

6. The computation of free enthalpy: long-range contributions

In the previous example of the change $Ne \rightarrow Na^+$ a full charge was created. Since the Coulomb interaction is inversely proportional to the first power of the charge-charge distance ($\sim r^{-1}$), the free enthalpy ΔG of creating a charge

will be dependent on the range (or cut-off radius R_c) of the interaction that is taken into account. This is illustrated in table VI, the larger R_c the lower ΔG will be. Even when the Born formula (Born, 1920; Straatsma 1987)

$$\Delta G = - \frac{e^2}{8\pi\epsilon_0} \frac{1}{R_c} [1 - 1/\epsilon] \quad (6.1)$$

is used to calculate the contribution from the continuum dielectric beyond R_c (with $\epsilon=80$ for water), the resulting ΔG value is sensitive to the R_c -value. Straatsma (1987) also showed that a variety of free enthalpy estimates can be obtained by using different values for the cut-off radii of solute-water and water-water interactions. When creating or annihilating a charge, the free energy change is very sensitive for the cut-off radius R_c that is used in the calculation.

Table VI
Free enthalpy as a function of cut-off radius^{a)}

Range of interaction R_c (nm)		ΔG (kJ/mol)		
solute-water	water-water	MD	Born correction interval (R_c, ∞)	Sum
0.9	0.9	-424	-76	-500
1.2	0.9	-461	-57	-518
0.9	1.2	-404	-76	-480
1.2	1.2	-429	-57	-486

^{a)}Data for the process of changing Ne (state A) into Na^+ (state B) in aqueous solution (512 H_2O , $\tau_{MD}=40\text{ps}$, $\Delta t=1\text{fs}$), taken from Straatsma (1987).

7. The computation of free enthalpy: dependence on model parameters

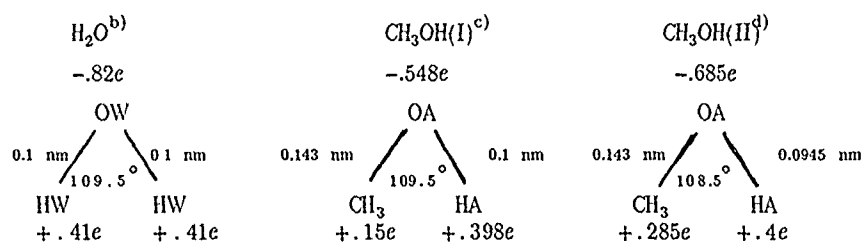
Even when no full charges are created the free energy of hydration can be very sensitive to model parameters that are used in the calculation. This is illustrated in tables VII and VIII for the simple process of changing a H_2O molecule into a CH_3OH molecule in aqueous solution. Model I and II have been taken from the literature (Hermans et al., 1984; Van Gunsteren & Berendsen, 1987; Jorgensen, 1981) and are both reasonable CH_3OH models, which however

differ by 7 kJ mol^{-1} in free enthalpy of hydration with respect to the SPC water model (Berendsen et al., 1981). When combining the geometry and Van der Waals parameters of model I with the charges of model II, a completely different ΔG value is found. This shows that the free enthalpy of hydration is very sensitive to the parameters used in the molecular model.

Since a free enthalpy estimate is critically dependent on sampling, cut-off radius and model parameters, it is easily possible that good agreement of simulated with experimental free enthalpy is due to compensation of errors.

Table VII
Models for methanol-water interaction^{a)}

Geometries and charges



Van der Waals parameters^{e)}

Atom pair	$C_{12} \text{ (kcal mol}^{-1} \text{\AA}^{12})$	$C_6 \text{ (kcal mol}^{-1} \text{\AA}^6)$
OW - OW	793.3 * 793.3	25.01 * 25.01
OW - CH ₃ (I)	421.0 * 2500	25.01 * 46.06
OW - OA (I)	793.3 * 600.0	25.01 * 23.25
OW - CH ₃ (II)	421.0 * 2820	25.01 * 48.99
OW - OA (II)	793.3 * 717.6	25.01 * 24.49

^{a)}Data from (Berendsen et al., 1981^{b)}; Hermans et al., 1984; Van Gunsteren & Berendsen, 1987^{c)}; Jorgensen, 1981^{d)})

^{e)}The hydrogen atoms do not have Van der Waals interactions.

Table VIII
Free enthalpy of methanol hydration

Process ^{a)}	ΔG (kJ/mol)
$H_2O \rightarrow CH_3OH$ (I)	7
$H_2O \rightarrow CH_3OH$ (II)	14
$H_2O \rightarrow CH_3OH$ (I,II) ^{b)}	0

^{a)}Change carried out over 20 ps, $R_c=0.8nm$, $\Delta t=2fs$, for a periodic box containing 216 H_2O molecules. The experimental value is 5 kJ mol^{-1} (Ben-Naim & Marcus, 1984)

^{b)}Geometry and Van der Waals parameters from model I, charges from model II (P.A. Kollman, private communication).

8. Spatial molecular structure determination by 2D-NMR

The last example of obtaining good agreement between theoretical and experimental data by compensation of errors lies in the area of structure determination by nuclear magnetic resonance (NMR) measurements (Wüthrich, 1986; Kaptein et al., 1988). High-resolution NMR is able to resolve individual proton resonances of proteins in solution. Once the observed resonances have been assigned to individual protons, two-dimensional (2D) nuclear Overhauser enhancement (NOE) spectra can be used to obtain upper (and lower) limits or distance constraints to the distances between pairs of protons. The next problem is the derivation of a 3D structure satisfying these distance constraints. Crude molecular structures that approximately satisfy the constraints can be refined by MD simulation using a simple energy term (Kaptein et al., 1985)

$$V_{dc} = \sum_{n=1}^{N_{dc}} \frac{1}{2} K_{dc} (r_{ij} - r_{ij}^n)^2 \quad (8.1)$$

which represents the set of N_{dc} distance constraints, where the distance constraint between atoms i and j is denoted by r_{ij}^n . The term V_{dc} is added to the normal interaction function V . By performing MD simulation and subsequent energy minimization (EM) one searches for a conformation which has a low total energy $E + E_{dc}$.

The term V_{dc} forces the molecule to satisfy the experimental data (constraints). Therefore, the larger its weight K_{dc} is, the smaller the distance constraint violations will be. But, the agreement with the experimental data is generally paid for by a higher energy E of the molecule, that is, the molecule adopts a more strained conformation in order to satisfy the distance constraints.

An example is found in table IX, where a short MD refinement of two lac repressor structures is shown. The starting structures were taken from (De Vlieg et al., 1986): the final (best) structure and a structure which still had its loop wrongly folded (after 20 ps of refinement). MD refinement with a large weight $K_{dc} = 17000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ makes the wrongly folded structure have lower constraint violations (0.180 nm) than the correctly folded structure (0.258 nm), but this occurs at the expense of a higher energy of $-2996 \text{ kJ mol}^{-1}$ compared to $-3053 \text{ kJ mol}^{-1}$. When the constraint energy term is switched off ($K_{dc}=0$), EM and MD do relax the molecule at the expense of larger constraint violations. We conclude that 3D structure obtained from NMR should not only show small violations of distance constraints, but must also have a low energy. Otherwise one may have obtained a wrongly folded structure by which the experimental data are satisfied at the expense of a strained non-stable conformation.

Table IX
MD refinement of lac repressor structures

	structure I loop wrongly folded			structure II loop correctly folded		
	constraint violation (nm)		energy	constraint violation (nm)		energy
	sum	average	(kJ/mol)	sum	average	(kJ/mol)
1. initial structure ^{a)}	3.242	0.015	-2115	0.405	0.002	-3092
2. after 5 ps MD +EM $K_{dc}=17000 \text{ kJmol}^{-1} \text{ nm}^{-2}$	0.180	0.001	-2996	0.258	0.001	-3053
3. after EM $K_{dc}=0$	0.523	0.002	-3083	0.461	0.002	-3100
4. after 5 ps MD + EM $K_{dc}=0$	3.297	0.015	-3032	1.823	0.008	-3102

^{a)}Initial structures were taken from De Vlieg et al. (1986).

Table X
Energies of X-ray and NMR protein structures

	number of residues	source of data or structure	N_{dc}	K_{dc}	average violation (nm)	energy ($\text{kJmol}^{-1}\text{nm}^{-2}$)	force field
aPP	36	X-ray ^{a)}				-2180	GROMOS ^{j)}
crambin	46	X-ray ^{a)}				-2161	GROMOS
BPTI	58	X-ray ^{a)}				-3529	GROMOS
phospholipase A2 123		X-ray ^{a)}				-7848	GROMOS
lac repressor	51	NMR ^{b)}	215	4000	0.003	-3091	GROMOS
tendamistat	74	NMR ^{c)}	842			-3140, -2834	GROMOS
crambin	46	X-ray/NMR ^{d)}	240	5000	0.033	-2247	CHARMM ^{k)}
hirudin	56	NMR ^{e)}	359	17000	0.016	-1138	CHARMM
histone H5	79	NMR ^{f)}	307	17000	0.015	-1527	CHARMM
CPI	39	NMR ^{g)}	309	33000	0.007	-724	CHARMM
phoratoxin	46	NMR ^{h)}	331	33000	0.010	-1029	CHARMM
α 1-purothionin	46	NMR ⁱ⁾	310	33000	0.023	-498	CHARMM

a) X-ray structures from Brookhaven Protein Data Bank (Bernstein et al., 1977)

b) De Vlieg et al. (1986)

c) Kline et al. (1988), lowest and highest energy of a set of structures is given

d) Clore et al. (1986a)

e) Clore et al. (1987a)

f) Clore et al. (1987b)

g) Clore et al. (1987d)

h) Clore et al. (1987c)

i) Clore et al. (1986b)

j) Van Gunsteren & Berendsen (1987)

k) Brooks et al. (1983)

How low should the molecular energy be? In table X we list the energies of a number of proteins for which a high-resolution X-ray structure is known. As expected, the energy depends on the size (number of residues) of the protein. The NMR based structures of the lac repressor and tendamistat both have energies comparable to those of X-ray determined structures. In contrast, the 5 last structures in table X display a relatively high energy, at least about 1000 kJ mol^{-1} higher than an X-ray structure of comparable size. This may be due to the fact that the use of a large K_{dc} will lower the constraint violations at the expense of a large internal energy. The discrepancy in energy cannot be due to the application of different force fields. CHARMM (Brooks et al., 1983) and GROMOS (Van Gunsteren & Berendsen, 1987) consist of comparable atom-atom interaction terms: MD refinement of crambin using theoretical distance constraints derived from the X-ray structure leads to an energy of $-2247 \text{ kJ mol}^{-1}$ in the CHARMM force field (Clore et al., 1986a). The GROMOS force field yields a comparable value of $-2161 \text{ kJ mol}^{-1}$ for the crambin X-ray structure.

We conclude that a wrongly folded structure may be forced to agree with a set of experimental distance constraints by using large weights in the distance constraint energy term. When a protein structure that is refined displays a relatively large energy - compared to the energies of X-ray structures of proteins of comparable size - this must be taken as a warning that a (partially) wrongly folded structure is obtained.

9. Conclusions

In this paper we have shown that good agreement between theoretically calculated and experimentally measured data does not necessarily imply that the theoretical model is correct. Compensation of errors may be at the basis of the good agreement. Especially when high-dimensional systems modelled with many parameters are concerned, it is relatively easy to choose or fit parameters such that good agreement is obtained for a limited number of observable quantities.

Acknowledgements

This work was supported in part by the Foundation for Chemical Research

(SON) and in part by a NATO Science Fellowship, both under auspices of the Netherlands Research Organisation (NWO). The hospitality of the Department of Pharmaceutical Chemistry of the School of Pharmacy of the University of California at San Francisco is gratefully acknowledged.

References

- Aqvist, J., Van Gunsteren, W.F., Leijonmarck, M. & Tapia, O. (1985). *J. Mol. Biol.* 183, 461-477.
- Beeman, D. (1976). *J. Comput. Phys.* 20, 130-139.
- Ben-Naim, A. & Marcus, Y. (1984). *J. Chem. Phys.* 81, 2016-2027.
- Berendsen, H.J.C. Postma, J.P.M., Van Gunsteren, W.F. & Hermans, J. (1981) In: *Intermolecular Forces*, B. Pullman ed., Reidel, Dordrecht, The Netherlands, pp. 331-342.
- Berendsen, H.J.C., Postma, J.P.M., Van Gunsteren, W.F., DiNola, A. & Haak, J.R. (1984). *J. Chem. Phys.* 81, 3684-3690.
- Berendsen, H.J.C. & Van Gunsteren, W.F. (1986). In: *Molecular-Dynamics Simulation of Statistical-Mechanical Systems*, Proceedings of the International School of Physics "Enrico Fermi", Course 97, G. Ciccotti & W.G. Hoover eds., North-Holland, Amsterdam, pp. 496-519.
- Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F. Jr., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *J. Mol. Biol.* 112, 535-542.
- Born, M. (1920) *Z. Phys.* 1, 45.
- Brooks, B.R., Bruccoleri, R.E., Olafson B.D., States, D.J., Swaminathan, S. & Karplus, M. (1983). *J. Comp. Chem.* 4, 187-217.
- Clore, G.M., Brünger, A.T., Karplus, M. & Gronenborn, A.M. (1986a). *J. Mol. Biol.* 191, 523-551.
- Clore, G.M., Nilges, M., Sukumaran, D.K., Brünger, A.T., Karplus, M. & Gronenborn, M. (1986b). *EMBO J.* 5, 2729-2735.
- Clore, G.M., Sukumaran, D.K., Nigles, M., Zarbock, J. & Gronenborn, A.M. (1987a). *EMBO J.* 6, 529-537.
- Clore, G.M., Gronenborn, A.M., Nilges, M., Sukumaran, D.K. & Zarbock, J. (1987b). *EMBO J.* 6, 1833-1842.
- Clore, G.M., Sukumaran, D.K., Nilges, M. & Gronenborn, A.M. (1987c). *Biochemistry* 26, 1732-1745.

- Clore, G.M., Gronenborn, A.M., Nilges, M. & Ryan, C.A. (1987d). *Biochemistry* 26, 8012-8023.
- Hermans, J., Berendsen, H.J.C., Van Gunsteren, W.F. & Postma, J.P.M. (1984). *Biopolymers* 23, 1513-1518.
- Jorgensen, W.L. (1981). *J. Am. Chem. Soc.* 103, 335-340.
- Kaptein, R., Zuiderweg, E.R.P., Scheek, R.M., Boelens, R. & Van Gunsteren, W.F. (1985). *J. Mol. Biol.* 182, 179-182.
- Kaptein, R., Boelens, R., Scheek, R.M. & Van Gunsteren, W.F. (1988). *Biochemistry* 27, 5389-5395.
- Kline, A.D., Braun, W. & Wüthrich, K. (1988). *J. Mol. Biol.* 204, 675-724.
- Levitt, M. (1983). *J. Mol. Biol.* 168, 595-620.
- Love, R.A. & Stroud, R.M. (1986). *Protein Engineering* 1, 37-46.
- Rijckaert, J.-P., Ciccotti, G. & Berendsen, H.J.C. (1977). *J. Comput. Phys.* 23, 327-341.
- Straatsma, T.P. (1987). Free energy evaluation by molecular dynamics simulations, thesis, University of Groningen.
- Van Gunsteren, W.F. & Berendsen, H.J.C. (1985). In: *Molecular Dynamics and Protein Structure*, J. Hermans ed., Polycrystal Book Service, P.O. Box 27. Western Springs, pp. 5-14.
- Van Gunsteren, W.F. & Berendsen, H.J.C. (1987) Groningen Molecular Simulation (GROMOS) library manual, Biomos, Nijenborgh 16, Groningen, The Netherlands.
- Van Gunsteren, W.F. & Weiner, P.K. eds. (1989), *Computation of Free Energy for Biomolecular Systems*, Escom Science Publishers, Leiden, The Netherlands.
- Verdegaaal, C.H.M., De Kok, A.J., Westerink, H.P., Van Boom, J.H. & Romers, C. (1981). *Acta Cryst.* B37, 1924-1926.
- Verlet, L. (1967). *Phys. Rev.* 159, 98-103.
- Vlieg, J. de, Boelens, R. Scheek, R.M., Kaptein, R. & Van Gunsteren, W.F. (1986). *Isr. J. Chem.* 27, 181-188.
- Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.

THEORETICAL STUDY OF A CONFORMATIONAL CHANGE IN AN ENZYME : METHODOLOGY AND
FIRST RESULTS ON CITRATE SYNTHASE

J. DURUP

Laboratoire de Physique Quantique (UA 505 du C.N.R.S.) 118, route de Narbonne
31062 Toulouse (France)

Following the lines and using the codes developed in Harvard by M. Karplus and his colleagues, we introduced an alternative strategy for dynamical calculations on proteins at various time scales. It essentially consists of the following steps.

(i) We constructed a set of internal vectors by the principle of the binary tree, adapted to the peculiar structure of each residue and to the secondary and tertiary structures of the protein. The transformation from Cartesian coordinates, velocities and energy gradients to the new set and vice versa needs purely topological, constant matrices, the operation of which requires very little computer time ; the kinetic energy operator remains diagonal and thus the hamiltonian dynamics reduces to simple Newtonian dynamics.

(ii) From these internal vectors we generated a set of relative polar coordinates, where each vector is referred to the immediately lower ones in the tree ; then we determined through a dynamical test the frequencies associated with each of these coordinate, which permits a classification of slow and fast modes.

(iii) We compared a regular dynamics by the CHARMM programs, using a 1 fs integration step, with a constrained dynamics using 8 fs steps after freezing most of the modes of frequencies larger than $ca\ 1.4 \times 10^{13}\ s^{-1}$, and affecting the other fast modes with a damping coefficient. It appears that the main features of the dynamics of the slower modes are preserved, apart from phase shifts occurring during the freezing process. These tests were performed on the dimeric enzyme citrate synthase (437 residues per monomeric unit).

(iv) Iteration of this procedure to larger time scales is possible and will soon be performed. One may then hope to obtain significant collective modes by diagonalization of the correlation matrix of slow mode amplitudes.

Details of this work will be submitted for publication in the Journal of Physical Chemistry.

DISCUSSION

PERAHIA - How do you explain that you obtain larger rms) deviations than conventional dynamics for some groups of the protein ? I ask this question because the dynamics using holonomic constraints give smaller rms values.

DURUP - For money-saving reasons we did not perform the necessary 30-picosecond (at least) pre-thermalization, but only 2 picoseconds, with the exact dynamics. This results in spurious drifts of some α -carbon atoms, as well in the exact as in the constrained dynamics, which exceed the actual fluctuation rms's.

ANGYAN - Do you see any special problems to include solvent modes in your procedure ?

DURUP - I expect to include in my programs solvent effect calculations, either through a collaboration with eastern-Europe colleagues (R. Zahradnik, F. Aschenbach) or using other authors' codes. This will require, as stated in my introduction, entropy as well as energy terms.

VERGOTEN - You are freezing modes with frequencies higher than 10^{13} s^{-1} i.e. for spectroscopist 330 cm^{-1} . Experimentally the $-\text{NH}_3^+$ or $-\text{CH}_3$ torsional modes appear in the range $200\text{-}250 \text{ cm}^{-1}$ that means they are not frozen in your procedure ?

DURUP - The 10^{13} s^{-1} figure I gave is an order of magnitude. The actual separation limit between faster and slower modes was rather $1.4 \times 10^{13} \text{ s}^{-1}$. Anyway it is not a strict separation because of the requirement that all identical modes of residues of the same kind have to be treated on an equal footing whatever their individual frequencies in their specific environments, for the sake of transferability of the code to any protein.

SQUMPASIS - Did somebody try to eliminate the fast motions analytically i.e. use the normal mode collective excitations instead of the cartesian coordinates and work with mode-mode coupling theories ?

DURUP - The answer is "yes" and references were given in my fourth viewgraph (B. Brooks and Karplus 1983, Levitt et al 1985, Noguti and Gô 1985, Tobias and C. Brooks 1988, Sanejouand and Perahia to be published).

NONARCHIMEDEAN MODELLING FOR BIOSYSTEMS EVOLVING ON MANY TIME SCALES

O. IORDACHE¹ and P.T. FRANGOPOL²

¹Polytechnic Institute of Bucharest, Department of Chemical Engineering,
1 Polizu Street, R-78126 Bucharest 12 (Romania)

²Institute of Physics and Nuclear Engineering, P.O.Box MG-6, R-76900
Magurele-Bucharest (Romania)

SUMMARY

The nonarchimedean (NA) analysis was used to describe relaxation processes in chemical compounds-membrane interactions and the selfexcitation of membranes. The method relies upon the replacement of the real time by elements of NA fields. Thus an enriched mathematical and physical structure results.

Particularly, it is shown that an asymptotic expansion formalism can easily accommodate nearly-exponential relaxations in multi-compartmental systems (e.g. the action potential of the frog sciatic nerve in the presence of cross-linking aldehydes).

Intermittent regimes in excitable membranes are studied using dyadic expansions and Walsh-Fourier analysis.

INTRODUCTION

The existence of more than one characteristic time scale is a well known feature of far-from equilibrium biosystems. For instance, a variety of distinct time scales is displayed by the ionic pumps considered now as channels whose energy barrier profile is transiently modified by the phosphorylation .-dephosphorylation cycle (refs. 1, 2).

A lot of time scales characterize the relaxation of disordered media, similar to membranes (ref. 3). Also, in complex biochemical reactions one or several steps of the reaction sequence become very rapid as compared to others and a chemical instability corresponding to a new time scale usually appears.

All such phenomena have been studied using the concept of NA time. The proposed NA formalism proves to be an appropriate modelling tool if the studied system can be divided into a number of hierarchised interacting subsystems. Such formalisms have been used to describe either nearly-linear phenomena (a perturbation method) or extremely nonlinear interactions (a boolean method).

This work is part of a recent attempt to apply NA analysis (see ref. 4) in chemistry and biophysics (refs. 5-8).

EFFECT OF PROTEIN CROSS-LINKING ALDEHYDES ON NERVE ACTIVITY

A multi-compartmental model

The necessity of multi-compartmental models of relaxation in the study of chemical compounds - membrane interactions is imposed by the fact that membrane excitability could depend on the collective behaviour of a number of molecular units (compartments) which all have to remain intact for the survival of the excitability.

Consider a system of $n+1$ identical compartments connected in series. Denote by $y(t)$ the time evolving physical variable of the relaxation process (in this case, the amplitude of the compound action potential of frog sciatic nerve (refs. 9-10)). The variable corresponding to the i -th compartment is $y^{(i)}(t)$.

The relaxation process of such system is described by:

$$\frac{dy^{(i)}(t)}{dt} = -y^{(i)}(t) + y^{(i-1)}(t), \quad 1 \leq i \leq n \quad (1)$$

where $t = at$ is the dimensionless time, τ is the time, and a is the decay or relaxation rate assumed to be the same for any compartment. The index $i = 0$ refers to the single compartment model (ref. 6). The solution of (1) with initial conditions:

$$y^{(0)}(0) = 1, y^{(1)}(0) = \dots = y^{(n)}(0) = 0 \quad (2)$$

is:

$$y^{(i)}(t) = t^i \exp(-t)/i! \quad (3)$$

To take into account the existence of different time scales we replace the time t by the expansion T :

$$T = t + \epsilon w_1 t + \dots + \epsilon^M w_M t \quad (4)$$

Here M is the number of time scales, ϵ is an expansion parameter and $w_m, m = 1, \dots, M$ are constants. Denote also $T = [t, w_1 t, \dots, w_M t]$. The use of (4) induces a similar structure of $y^{(i)}(t)$, that is:

$$Y^{(i)}(T) = y_0^{(i)}(t) + \epsilon y_1^{(i)}(t) + \dots + \epsilon^M y_M^{(i)}(t) \quad (5)$$

where $y_m^{(i)}(t)$ are real functions.

Both (4) and (5) can be viewed as elements of an NA field (ref. 4). This NA field is endowed with addition and multiplication rules defined as follows:

if $A = [a_0, a_1]$ and $B = [b_0, b_1]$ then $A + B = [a_0 + b_0, a_1 + b_1]$

$$AB = [a_0 b_0, a_0 b_1 + a_1 b_0].$$

In the new frame the relaxation will be described by the system of NA equations:

$$\frac{dy^{(i)}}{dT} = -y^{(i)} + y^{(i-1)} \quad (6)$$

The solution of (6) is:

$$y^{(i)}(T) = CT^i \exp(-T)/i! \quad (7)$$

where C is an NA constant. We take $w_m = 1$ and consequently $C = [1, 1, \dots, 1]$ (a normalization condition). Expanding the exponential from (7) in power series of T and using the operations in the NA field, one obtains:

$$y^{(i)}(T) = t^i \exp(-t)/i! \sum_{m=0}^M \epsilon^m L_m^{(i)}(t) \quad (8)$$

where $L_m^{(i)}(t)$ are the Laguerre polynomials:

$$L_m^{(i)}(t) = \sum_{k=0}^m \binom{m+i}{m-k} (-t)^k / k! \quad (9)$$

A general, calculated solution of (6) valid for different w_m and initial condition is:

$$y_c^{(i)}(t) = t^i \exp(-t)/i! \sum_{m=0}^M q_m^{(i)} L_m^{(i)}(t) \quad (10)$$

The coefficients $q_m^{(i)}$ are related to the contribution of the m-th scale of the process.

In the NA formalism we obtain the minimum (null) distance between $y_c^{(i)}(t)$ given by (10) and experimental values $y^{(i)}(t)$ if:

$$\binom{m+i}{m} q_m^{(i)} = \int_0^\infty y^{(i)}(t) L_m^{(i)}(t) dt \quad (11)$$

Note that in the NA frame we cannot discriminate between the solution corresponding to different M (ref. 5). A truncation number M in (10) that is the best model is selected using the sum of squares of deviations:

$$S(M) = \int_0^\infty (y_c^{(i)}(t) - y^{(i)}(t))^2 dt \quad (12)$$

Analysis of experimental data

The predictions of the NA model are used to describe the effect of cross-linking aldehydes on the amplitude of the compound action potential of frog sciatic nerve (formaldehyde f.a., 0.10%, crotonic aldehyde c.a., 0.10% and glutaraldehyde g.a. 0.25%) (refs.9, 10).

The relaxation rate a is obtained taking $a = (n+1)/\bar{\tau}$. Here $\bar{\tau}$ is the mean relaxation time of the studied system (45.85 min for f.a., 6.65 min for c.a. and 76.80 min for g.a.).

Using (11) with $y^{(i)}(t)$ obtained from experiments one gets $q_m^{(i)}$ and the objective function $S(M)$ for different numbers of compartments, $n+1$, and of scales, $M+1$.

Table 1 contains the values of $S(M)$, at $n = 1$, for different M . When M increases from 0 to 3 the sum of squares of deviations decreases sharply, while for values of M above 3 it remains practically constant. It has been concluded that two compartments and four scales of time provide an optimal fitting of the relaxation process for all cross-linking aldehydes.

TABLE 1
Objective functions $S(M)$

M	0	1	2	3	4	5
f.a.	4.33	.37	.22	.11	.10	.09
c.a.	.164	.044	.023	.013	.013	.012
g.a.	9.82	.21	.07	.026	.028	.025

The number $n+1$ of compartments appearing in the model could be viewed as the number of targets which must be "hit" by aldehydes in order to block impulse propagation in each nerve fibre. In this model the time scales correspond to parallel pathways with different permeabilities. This interpretation agrees with the existence of four repeated regions or organizational levels in the sodium channels (refs. 1,2,6,7).

The fact that we obtained four time scales, two compartments and comparable coefficients $q_k^{(1)}$ (see Table 2) for all cross-linking aldehydes suggests that a similar mechanism of inactivation takes place in every case.

TABLE 2
Coefficients $q_k^{(1)}$

k	0	1	2	3
f.a.	.98	.054	-.155	-.115
c.a.	.98	.090	-.126	-.092
g.a.	.99	.033	-.194	-.150

FLUCTUATIONS IN MEMBRANES

Various biosystems exhibit electrical oscillations over bundles of cells in nerve and muscle tissues, intermittent bursting or beating domains of behaviour. Similar phenomena have been observed in neural networks and in artificial membrane. For instance, quasi-periodicity of relaxation type in the interface electric potential of an oil/water system in the presence of surfactants has been reported (ref. 11). Self-excitation and intermittences also characterize porous doped membranes (ref. 12). We refer, in the following, to such porous membranes due to their similarity with biomembranes.

A simple model governing the mass transport through a pore is:

$$\frac{dy}{dt} = -ky \quad (13)$$

where $y(t)$ is the concentration, k is the transport coefficient, and t is the time.

Such a model cannot explain the intermittent behaviour observed in natural and artificial membranes.

An approach based on an NA formalism is now considered. We suppose that, due to extremely non-linear interactions characterizing the transmembrane transport, a self-organization of movements, that is a hierarchy of roll-cells as shown in Fig. 1, appears in the doped pore. Different levels in this hierarchy are indexed by $m = 0, 1$ and 2 . Making a choice of the unit time we consider that the transport time is 2^m for a cell of the type m .

On account of the mechanism of transport through such structured pores, the time t and the coefficient of transport k are considered as elements of a 2-series field (ref. 13). Denote these elements by T and K , respectively. We have $T = \sum_j t_j 2^j$ and $K = \sum_j k_j 2^j$ where j is a natural number and t_j and k_j are the digits 0, 1.

Observe that a hierarchical tree as shown in Fig. 2 could be associated in a one-to-one correspondence to the hierarchy of roll-cells. The lower ends of the branches of this tree represent the possible times (000, 100, 010 and so on). They are denoted using precisely the dyadic notation (for instance the time 011 signifies $0.2^0 + 1.2^1 + 1.2^2$).

The dyadic expansions give detailed descriptions of the time and of the coefficient of transport by cellular motions. The addition " \oplus " and the multiplication " \otimes " in this 2-series field will be defined by:

$$T \oplus K = \sum_j (t_j + k_j)_{\text{mod} 2} 2^j, \quad T \otimes K = \left(\sum_j t_j k_j \right)_{\text{mod} 2} 2^j \quad (14)$$

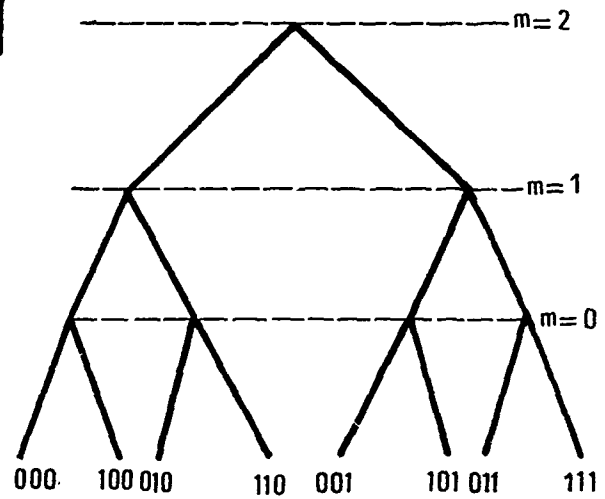
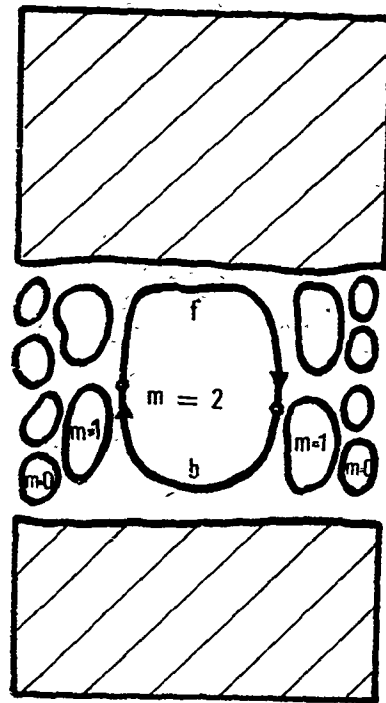


Fig.1. Structured pore

Fig.2. Associated tree

TABLE 3

$$y(T) = \exp(-K \otimes T), \quad 0 \leq K, \quad T \leq 7$$

$K \backslash T$	0	1	2	3	4	5	6	7
0	1	1	1	1	1	1	1	1
1	1	1/e	1	1/e	1	1/e	1	1/e
2	1	1	1/e	1/e	1	1	1/e	1/e
3	1	1/e	1/e	1	1	1/e	1/e	1
4	1	1	1	1	1/e	1/e	1/e	1/e
5	1	1/e	1	1/e	1/e	1	1/e	1
6	1	1	1/e	1/e	1/e	1/e	1	1
7	1	1/e	1/e	1	1/e	1	1	1/e

The addition rule is justified by the fact that two identical steps (at the same level in the hierarchy) should correspond to the situation in which no transport takes place that is to the null time. For instance the forward step 001, denoted by f in Fig. 1, followed by the backward step 001, denoted by b , gives according to (14) $001 \oplus 001 = 000$.

The multiplication rule " \otimes " is in fact a natural rule of coupling the transport potentialities k_j and durations t_j .

The solution of the model (13) is in this 2-series field:

$$y(T) = \exp(-K \otimes T) \quad (15)$$

Table 3 presents such solutions for $0 \leq K, T \leq 7$. Notice that by simple linear transformations the solutions reduce to the well-known Walsh functions (ref. 14) defined by:

$$W_M(K, T) = (-1)^{\sum_{j=0}^{m-1} k_j t_j} \quad (16)$$

with $K = k_0 2^0 + \dots + k_{m-1} 2^{m-1}$, $T = t_0 2^0 + \dots + t_{m-1} 2^{m-1}$, $M = 2^m$ (in Table 3, $M = 8$). Consequently, the general calculated solution of (13) is written as:

$$y_c(T) = \sum_{K=0}^{M-1} q_K W_M(K, T) \quad (17)$$

The Walsh-Fourier coefficients q_K depend on the contribution of the sequence K (see ref. 14). They could be obtained from:

$$q_K \sum_{T=0}^{M-1} W_M^2(K, T) = \sum_{T=0}^{M-1} y(T) W_M(K, T) \quad (18)$$

where $y(T)$ is the experimentally recorded fluctuation.

CONCLUSIONS

The coefficients q obtained in both NA models developed here could be interpreted in the form of a spectrum giving a measure of the contribution of different scales involved in the process.

The nearly-exponential and the intermittent relaxations studied here represent, in our opinion, suitable candidates for physically modelling the biological drug action, the neural networks and other biosystems of practical interest.

Note that the main point of the present analysis is that the time is considered to pertain to some NA frames. This allows to describe a large class of "exotic" relaxations using very simple models such as (1) and (13). However some of the manifestations of the NA methods are new and it is still difficult to obtain physical insight into them.

REFERENCES

- 1 L.S.Liebovitch, Analysis of fractal ion channel gating kinetics: kinetic rates, energy levels and activation energies, *Math.Biosci.*, 93 (1989) 97-115
- 2 P.Läuger, Internal motions in proteins and gating kinetics of ion channels, *Biophys J.*, 53 (1988) 877-884
- 3 M.F.Schlesinger, Fractal time in condensed matter, *Ann.Rev.Phys.Chem.* 30 (1988) 269-290
- 4 W.H.Schikhof, The ultrametric calculus, Cambridge University Press, Cambridge, 1984
- 5 O.Iordache, R.Isopescu, A.Isopescu, A.Iacobini and G.Jinescu, Non-archimedean models of mixing i, *Chem.Engng.Sci.*, 43 (1988) 693-697
- 6 O.Iordache and P.T.Frangopol, Relaxation in biophysical systems, *Rev.Roum.Phys.*, 33 (1988) 1159-1162
- 7 O.Iordache, M.Apostol and P.T.Frangopol, Non-exponential relaxation of drug-membrane interaction. Effects of cross-linking aldehydes and procaine. *Rev.Roum.Biochim.* 25 (1988) 275-281.
- 8 O.Iordache and P.T.Frangopol, Nonarchimedean formalism for entropy balance in biosystems, *Rev.Roum.Phys.*, 34 (1989) 217-227.
- 9 D.G.Margineanu, E.Katona and J.Popa, Effect of Protein cross-linking aldehydes on nerve activity, *Arch.Internat.Physiol.Biochim.*, 89 (1981) 159-165.
- 10 D.G.Margineanu, E.Katona, J.Popa, Kinetics of nerve impulse blocking by protein cross-linking aldehydes, *Biochim.Biophys.Acta*, 648 (1981) 581-586.
- 11 M.Dupeyrat and E.Nakache, Hydrodynamical instabilities of some interface as inferfacial membrane models; Correlation with electrical interfacial phenomena in E.Roux (Ed.), *Electrical phenomena at the biological membrane level*, Elsevier, Amsterdam, 1977, pp 377-388
- 12 K.Yoshikawa, K.Sakabe, Y.Matsubara, T.Ota, Self-excitation in a porous membrane doped with sorbitan monooleate (Span 80) induced by Na^+/K^+ concentration gradient, *Biophys.Chemistry*, 21 (1985) 33-39
- 13 M.H.Taibleson, Fourier analysis on local fields, *Mathematical Notes*, Princeton University Press, Princeton, 1975.
- 14 H.F.Harmuth, Sequency theory, Foundation and applications, Academic Press, New York, 1977.

MOLECULAR DYNAMICS SIMULATIONS OF A SOLVATED PROTEIN: ANALYSIS OF STRUCTURAL AND ELECTROSTATIC PROPERTIES

Shoshana J. Wodak¹, Daniël Van Belle², Matheus Froeyen¹, and Martine Prévost¹

¹ Unité de Conformation de Macromolécules Biologiques, Université Libre de
Bruxelles, CP160, P2, Ave. P. Héger, B-1050 Bruxelles (Belgium)

² Plant Genetic Systems, Université Libre de Bruxelles, CP160, P2, Ave. P. Héger,
B-1050 Bruxelles (Belgium)

SUMMARY

Molecular dynamics simulations of a small protein, barnase, in presence of explicit water molecules are described and results from a 120 ps trajectory of the system are analyzed.

The deviations of the average protein conformation from the starting crystal structure is small (-1.1 Å rms for backbone atoms), and the agreement between computed and crystallographic atomic fluctuations is satisfactory for portions of the protein that do not participate in crystal contacts. Other parameters such as the accessible area of the protein and its molecular volume change by no more than 5% and 2% respectively. The structure of water around polar and non-polar groups on the protein surface also seems reasonable in that it agrees well with previous observations made in hydrated crystals or in simulations of small solvated molecules.

The thrust of our study concerns the use of the generated trajectory to obtain a detailed microscopic description of electrostatic interactions in the protein. Contributions to these interactions from permanent protein dipoles, from orientable solvent dipoles, and from electronic polarizability are evaluated. Although the analysis must still be qualified as preliminary, a number of clear trends emerge. The contribution of water to local fields inside the protein is substantial. In exposed parts of the protein, it affects field magnitudes and field orientation respectively by about 60% and 43 degrees on the average. In buried regions, where residues expose less than 10% of their surface to solvent, this contribution is reduced to 12% (in field magnitude) and 20 degrees (in field orientation) where it becomes comparable to the contribution from electronic polarizability. The latter averages 13% in field magnitude and 12 degrees in field orientation, and is not sensitive to solvent exposure. Detailed analysis of both solvent and electronic induction effects, shows that they display an appreciable degree of inhomogeneity throughout the protein matrix, suggesting that their relative importance may vary quite dramatically according to the local environment.

INTRODUCTION

Electrostatic interactions are among the most important factors in determining the structure and function of proteins. They are involved in enzymic mechanisms (1), allosteric control (2), specific ligand binding (3-4) as well as in other essential phe-

nomena such as protein stability (5) and folding (6). It is therefore important to be able to evaluate correctly electrostatic energies and forces acting upon atoms using available X-ray structures of proteins or conformations generated by computer simulations. This is however, a difficult problem owing to the inherent complexity of these interactions for solvated macromolecules, and owing also to the long range of electrostatic forces relative to the forces associated with other interactions. The major bottleneck lies with the proper representation of dielectric or screening effects associated with the orientable permanent dipoles of surrounding solvent molecules and with the inherent many-body nature of electronic polarizability. In present computational procedures, two main approaches are being used to represent electrostatic interactions: a microscopic approach, usually incorporated in Molecular Mechanics and Molecular Dynamics simulations, in which electrostatic interactions for both the protein and solvent components are calculated by summing pairwise interactions between permanent dipoles and charges whose positions in space are known, and a macroscopic approach based on continuum electrostatic models and which requires solving the Poisson-Boltzmann equation. A microscopic approach in which a detailed representation of the protein atoms and explicit solvent molecules is associated with Molecular Dynamics simulations to describe the time dependent behaviour of the system, might in principle be expected to provide an adequate description of electrostatic properties. But, while this approach is potentially quite accurate, it is in practice computationally overwhelming. Due to the heavy computational requirements, many approximations have been introduced resulting mostly in hybrid models that incorporate mixed microscopic-macroscopic representations. In these, solvent effects are mimicked by distance dependent dielectric functions (7), or by Langevin type dipoles (8). In more recent studies, explicit solvent molecules have been included, but contributions from electronic polarizability have not (9-11). The macroscopic continuum approaches, on the other hand, involve the simplifying assumption that protein interiors are homogeneous low dielectric media, surrounded by solvent of high dielectric constant. First introduced by Tanford & Kirkwood in 1957 (12), they have been recently generalized to non spherical geometries with the implementation of numerical procedures for solving Poisson-Boltzmann Equation (13-15), to improve the treatment of solvation energies (16) and to include effects of ionic strength (17). Recent results obtained by these methods on evaluating changes in pKa of protein groups (18-19) and in redox potentials (20) have been encouraging. But the main limitations of the approach, in particular, the validity of the uniform dielectric model for the protein interior, and the precise influence of the non physical description of the protein/solvent interface, are still not well understood.

With constant increase in available computer power, the obstacles encountered in implementing detailed microscopic representations of large molecular systems are gradually receding. It becomes possible now to envisage extensive testing of such models in computational procedures and confront them with experimental data. In a recent study (21), we described the implementation of a microscopic model for evalu-

ating electrostatic interactions in proteins which incorporates electronic polarizability effects. Following the approach of a number of authors (22-23), isotropic polarizabilities have been assigned to individual atoms, and resulting departure from pairwise interactions have been treated by a self consistent iterative procedure. This procedure was incorporated in Molecular Mechanics calculations, and methods for its practical use in Molecular Dynamics simulations were proposed. Moreover, the contribution of electronic polarizability to electrostatic potential energies, but more particularly to local electrostatic fields and to dipoles moments of structural elements in the protein were thoroughly analyzed. The study did however not include contributions from solvent molecules, or from protein flexibility, since computations were performed only on static protein crystal structures in vacuum. Here, we extend the analysis to include both contributions, and present highlights of our initial results. Molecular Dynamics simulations of a small protein barnase in presence of explicit water molecules are used to generate molecular trajectories of the system. Although detailed analysis of the trajectories is still in progress, evidence is presented that it provides a very satisfactory description of the protein and water portions. The generated trajectories are then used to obtain a detailed microscopic description of electrostatic interactions in the protein. Contributions to these interactions from permanent protein dipoles, from orientable solvent dipoles, hereafter referred to as solvent polarization, and from electronic polarizability are analyzed.

MOLECULAR DYNAMICS SIMULATIONS OF BARNASE IN WATER

A time dependent trajectory of the protein barnase in presence of explicit solvent molecules is computed by solving Newton's equations of motion for each particle in the system, with forces evaluated as the negative gradient of a classical empirical potential energy function.

The simulations are performed using recently implemented vectorized procedures in the BRUGEL package (24) in the microcanonical ensemble (N, V, E), at room temperature ($\langle T \rangle = 304$ K). The force-field is derived from a recent version of the CHARMM potentials that includes explicit contributions from aliphatic hydrogens in the protein portion. The solvent is modeled by the three-center charge TIPS model (25) with atomic charges $q(O) = -0.814$ and $q(H) = +0.407$ esu, and van der Waals parameters $\epsilon = -0.1521$ kcal/mol and $\sigma = 3.1506$ Å. Electrostatic interactions are modeled by the usual Coulomb potential with a dielectric constant $\epsilon = 1$. Effects due to induced dipoles are not included in generating the trajectories but applied afterwards to computed conformations in a perturbation-like approach. The starting conformation of the system consisted of one of the three molecules (molecule C) in the asymmetric unit from the 2 Å resolution refined crystal structure of barnase (26), crystallographically determined water positions located within 4 Å of a protein atom, and randomly oriented water molecules placed on a cubic lattice in a rectangular box

(dimensions : 49.68 Å x 37.16 Å x 49.68 Å). The system contains 8777 atoms. These include 1700 protein atoms, comprising all hydrogen positions, generated using standard bond distances and angles (27), 94 water molecules positioned crystallographically, and 2265 generated waters. With the applied boundary conditions, atoms of proteins in adjacent boxes are separated by at least three water layers. Relevant simulation data and parameters are summarized in Table 1.

TABLE 1

Protein-Solvent Simulation conditions

Thermodynamic ensemble :	Microcanonical (N,V,E)
Integration algorithm :	Verlet (40),
Constraints* :	protein : bond distances water: bond distances and bond angles
Integration time-step:	0.002 ps
Long-range interactions:	7 Å Cutoff distance
Shifting Function:	MEI4**
Periodic boundary:	box dimensions: 49.68 X 37.16 X 49.68 Å
Thermalization:	(velocity rescale), 4 ps
Equilibration :	10 ps.

Simulations have been performed using the BRUGEL package (24) and code vectorized to run on the Convex-C1 computer. Constraints* are applied using the Shake procedure (41) and MEI4** is the shifting function previously tested using Integral Equation methods (28) (see text).

Particularly noteworthy is the chosen treatment of long-range (Coulomb) interactions. This treatment consists in applying a 7 Å radial atom-atom cutoff distance together with a modified Coulomb interaction active over the entire distance range:

$$U'(r) = U(r) \cdot S(r) \quad (1)$$

With U' and U being respectively the modified and unmodified Coulomb potential, and r the interatomic distance. $S(r)$ is the MEI4 shifting function (28) which has the following form:

$$S(r) = 1 - 2\left(\frac{r}{r_c}\right) + \left(\frac{r}{r_c}\right)^2 \quad r < r_c \quad (2)$$

$$= 0 \quad r \geq r_c$$

With r_c being the cutoff distance.

In a recent study (29), several truncation schemes for long-range interactions have been calibrated against the Ewald-Kornfeld-summation method (30) in simulations of pure liquid water. It has been shown that the MEI4-function performs best with three-center models such as SPC and TIPS, both with respect to structural and thermodynamic properties, hence the choice of MEI4 for the present protein-water simulations in which the water is expected to play an important role.

Analysis of simulation results

Results reported here are based on a 120 ps (1 ps = 10^{-12} s) trajectory of the system described above. According to usual criteria such as convergence of the average potential and kinetic energy components, size of the energy fluctuations (< 1%) and convergence in rms deviations from the crystal structure, the system appears to be at equilibrium.

Protein conformation and atomic fluctuations

The average conformation of the protein was computed from the 120 ps trajectory and compared to the starting crystal structure using coordinate superpositions (31). The computed rms deviations of backbone atoms is 1.13 Å and that of all the atoms in the structure (including aliphatic hydrogens) is 1.46 Å. This is the lowest value reported so far for simulations of solvated proteins where rms deviations of C α atoms usually range around 1.8–1.9 Å, except for the recently reported simulation by Levitt (11) where rms deviation of similar size (1.18 Å) have been obtained for non-hydrogen atoms of Bovine Pancreatic Trypsin Inhibitor in a 200 ps MD simulation. Preliminary results on a detailed comparison between the protein conformation averaged over the 120 ps trajectory and the crystal structure show that structural parameters such as backbone dihedral angles, and H-bonding interactions are in general very well conserved. The correspondence between the B-factors (also called temperature factors (32)) of barnase computed from the 120 ps MD simulation and those obtained from X-ray data is illustrated in Fig. 1. It is in general satisfactory, particularly in regions not involved in crystal contacts, where actual values of B-factors are often closely similar. Regions involved in crystal contacts display on the other hand important discrepancies in computed versus crystal B-factors, with the latter being always smaller. While increased flexibility in such regions can be expected to occur upon exposure to solvent, the large magnitude of some of the observed differences suggests that other phenomena, such as incomplete conformational relaxation in these regions, may be at play. Other parameters such as the protein accessible surface area and molecular volumes – computed using an analytic algorithm implemented in the BRUGEL package (Alard et al., in preparation) – have also been monitored along the trajectory (Fig. 2). After 120ps, the accessible surface area and the molecular volume of barnase show only a modest changes, respectively by -5% and -2 % (values based on comparison of the final spot structure with the crystal structure). While it may be premature to make definitive conclusions, our results suggest that the behavior of the

protein part in presence of water is very satisfactory. In order to assess if this should be attributed to artefacts in the force-field (which would act to curb substantial departures from the crystal structure) or to beneficial effect from adequately modeled surrounding solvent molecules, results from this simulation should be compared to those from vacuum simulations (presently in progress).

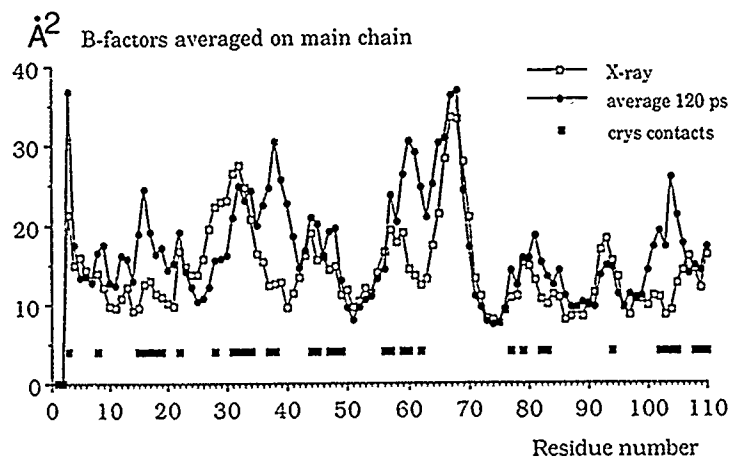


Fig 1: Comparison of B-factors of Ca atoms of barnase computed from the 120 ps simulation and from the refinement of the 2Å resolution X-ray structure of barnase. The ordinate shows B-factor magnitudes in units of \AA^2 . The abscissa shows residue numbers in barnase. Residues that bury 10 \AA^2 or more of surface area in an interaction with another molecule in the crystal of barnase are displayed.

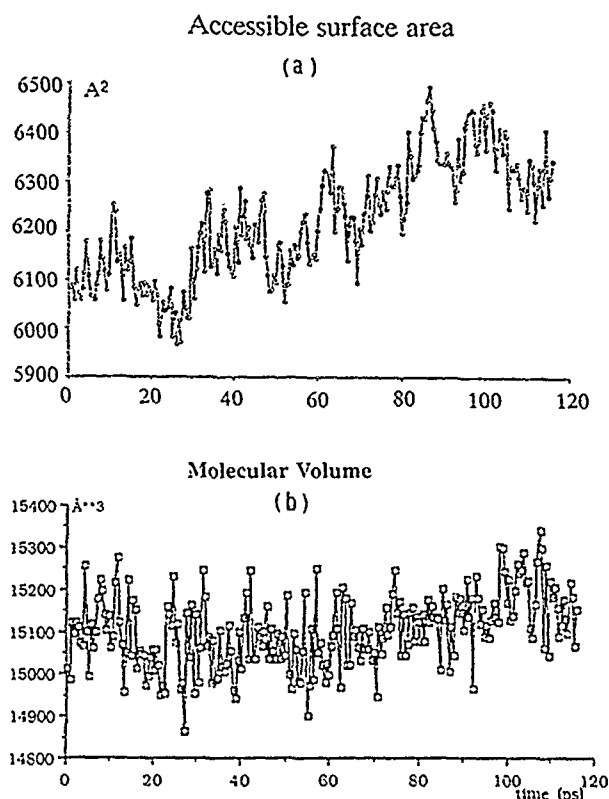


Fig 2: Accessible surface areas (a) and molecular volume (b) of barnase conformations along the 120 ps molecular dynamics trajectory. Accessible areas and molecular volumes have been computed using an analytic algorithm implemented in BRUGEL (Alard et al, in preparation).

Water structure around the protein

Molecular Dynamics simulations can provide a very accurate picture of the structure and dynamics of water molecules close to and in contact with the protein surface. A number of recent studies (9-11) are already contributing to change our view from that, provided by X-ray and Neutron diffraction studies, where hundreds of molecules are considered to be statically bound to the protein, to that in which water at the protein surface is generally not very different from bulk water, implying that many fewer water molecules are truly immobilized on the protein surface. Here we present preliminary results on the analysis of water structure at the surface of barnase as seen in our simulation. Radial distributions of water oxygens and hydrogens around specific protein atoms in non-polar, polar and charged groups, have been computed in exposed side-chains as determined from solvent accessibility calculations

(34). A sample of these distributions is shown in Fig. 3(a-d). Positions of maxima in the radial distribution functions involving polar and charged protein groups are within the expected H-bonding distances of the groups involved. A good example is the distribution of water oxygens around the oy group of asn-22 shown in Fig. 3a. The maximum of this distribution occurs at 2.63 Å, in good agreement with results from previous simulation (35). Further analysis of water structure around polar groups, shows that acceptor-proton distances are significantly shorter when the proton belongs to a water molecule than when it belongs to the protein as seen for example, from the peak positions of $g(r)$ involving the η OH group of Tyr-17 in barnase (Fig. 3(c-d)). Since we evaluate the Coulomb energy to be roughly the same in both cases this may be due to the lack of van der Waals parameters for the water hydrogens in the TIPS model used here. In addition we find that generally, proton-acceptor distances in our simulation are shorter than those found in small hydrated crystals (36), but the reasons for this are presently not clear. Radial distributions of water atoms around non-polar groups have also been analyzed. The peak in the water oxygen distribution around the non-polar methyl group of ala-32 in barnase (Fig. 3b) occurs at 3.5 Å, a somewhat shorter distance from those obtained previously in simulations of dilute solutions of alkanes and peptides: 3.6 Å and 3.7 Å, respectively for the methane-water first peak (37), and for both the butane-water (38) and peptide-methyl-water (35) first peaks.

A quite useful pictorial representation of water structure around specific protein side-chains can be obtained by representing water molecules from individual snapshots along the trajectory in a common local reference frame attached to the corresponding side-chains, examples of such representations were given during the oral presentation, but are not included here since they require colour. Further analysis of the structure of water at the protein surface, as well as a detailed study of the dynamic properties of water as a function of its distance to the protein, is in progress.

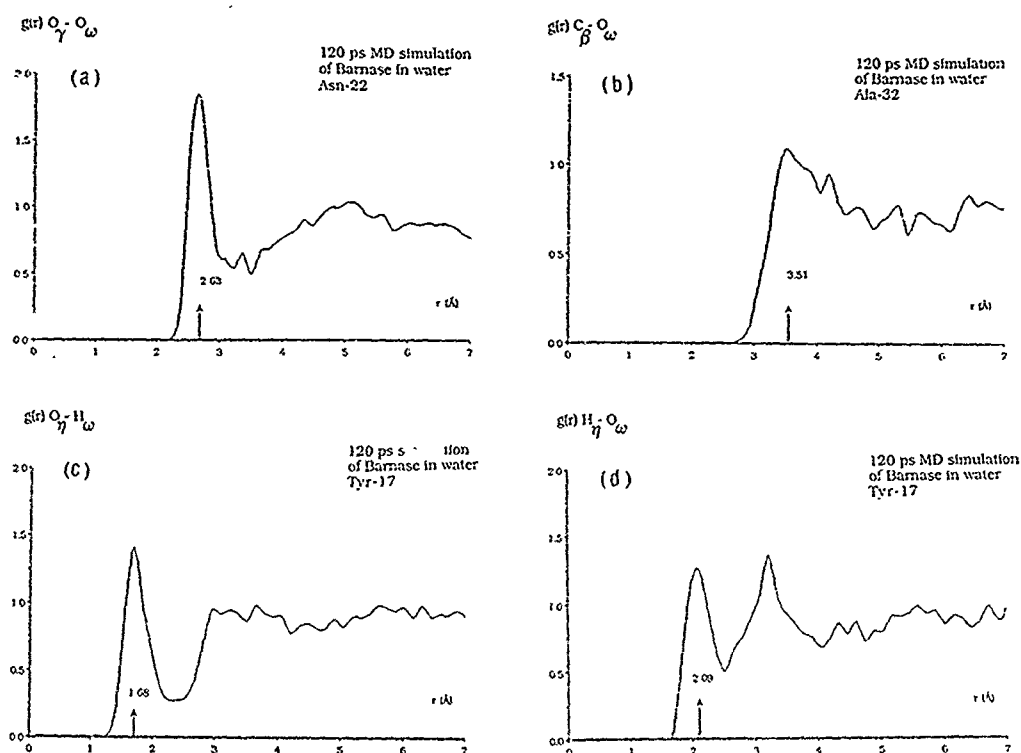


Fig 3: Radial distribution functions $g(r)$ for water atoms around solvent exposed groups of barnase.

(a) $g(r)$ of water oxygens (O_w) and the O_γ atom in the carbonyl group of Asn-22. (b) $g(r)$ of water oxygens and the C_β atom of Ala-32. (c) $g(r)$ of water hydrogens and the O_η group of Tyr-17, and (d) of water oxygens and the H_η group of tyr-17.

ELECTROSTATIC INTERACTIONS IN A SOLVATED FLEXIBLE PROTEIN

Contributions from electronic polarizability

Having convinced ourselves that the simulations yield, at least to a first approximation, a reasonable description of the protein and water portions, we proceed to analyze the contributions from electronic polarizability as a first step towards obtaining a detailed microscopic description of the different contributions to electrostatic interactions in solvated proteins, which is the main purpose of our study. The contributions of induced dipoles to electrostatic interactions in static protein crystal structures in vacuum have been analyzed previously (21). It has been shown that induced protein dipoles affect appreciably local electrostatic fields in magnitude and direction in a manner that is strongly influenced by the microscopic environment in the pro-

tein. But since effect of solvent and protein flexibility were not considered then, the possibility remained that induction effects and their inhomogeneity have been overestimated. This is one of the question we address here by analyzing the effects of protein and solvent dynamics on the contribution from induced dipoles to local fields inside the protein. For that purpose, averages and fluctuations of induced dipoles and of local fields at atomic positions are computed in conformations generated during the simulation. This perturbation-like approach is not a substitute to the correct but more time consuming procedure in which induced dipoles effects are included in the simulations. It is used here only as a first approximation to obtain indications on possible trends.

To calculate induced dipoles and their contributions to local fields we use procedures previously described (21) where isotropic polarizabilities are assigned to individual atoms and the resulting deviation from pairwise interactions is treated by a self consistent iterative procedure.

In this formalism, the total local field \vec{E}_i^t acting, say, on a protein atom i is expressed as follows:

$$\vec{E}_i^t = \vec{E}_i^q + \vec{E}_i^\mu \quad (3)$$

The first term \vec{E}_i^q , on the right hand side, is the electrostatic field on atom i due to all the partial charges (or permanent dipoles) of the system, excluding itself:

$$\vec{E}_i^q = \sum_{j \neq i} \frac{q_j}{r_{ij}^3} \vec{r}_{ij} \quad (4)$$

The second term in Eqn (3), \vec{E}_i^μ is the electrostatic field acting on the i th dipole (or atom) due to all the induced dipoles in the system excluding itself:

$$\vec{E}_i^\mu = - \sum_{j \neq i} \frac{1}{r_{ij}^3} (\vec{\mu}_j - 3(\vec{r}_{ij} \cdot \vec{\mu}_j) \frac{\vec{r}_{ij}}{r_{ij}^2}) \quad (5)$$

With $\vec{\mu}_j = \alpha_j \vec{E}_j$, α_j the scalar atomic polarizability, and r_{ij} the interatomic distance. Values of atomic polarizabilities are taken from the CHARMM library (39). Interactions between atoms separated by 2 covalent bonds or less are not considered.

Time averages and fluctuations of the relative orientations and sizes of induced fields \vec{E}^μ versus fields due to partial charges \vec{E}^q , are computed at atomic positions in protein conformations generated during the 120 ps trajectory. We find that the distributions of the atomic time averaged ratio of field magnitudes $\langle E^\mu / E^q \rangle$ remain very similar for the same trajectory of barnase in presence (Fig. 4a), and in absence (Fig. 4b) of surrounding water molecules (computed using the same distance cutoff

values and shifting function), with mean values for these ratios of 0.27 and 0.30 respectively. These distributions resemble moreover closely the one obtained for the ra-

tio E^μ/E^q in the static crystal structure of barnase in vacuum (Fig. 4c), where the mean is 0.31. This indicates that protein flexibility on the one hand, and presence of surrounding water molecules on the other do not contribute to average out electronic polarizability effects inside the protein, at least not at the time scale of the present simulation. This is further confirmed by the distributions of the atomic averages of the relative orientations between induced and permanent field vectors, shown in Fig. 5a-c. These distributions have been computed for barnase conformations along the same MD trajectory in presence and in absence of solvent, and for the static crystal structure in vacuum. All three are equally broad, and follow the shape of a random distribution. This however, should not be taken to imply that water has little influence on induced fields in individual locations in the protein. Indeed, we show below that water has an appreciable effect on local Coulomb fields (fields due to partial charges). This in turn affects induced dipoles. Fig. 6a shows the distribution of

atomic time averaged ratios $\langle E_{p+w}^\mu/E_p^\mu \rangle$, where E_p^μ , and E_{p+w}^μ are the magnitudes of local induced field respectively in absence and in presence of water molecules computed along the 120 ps trajectory. On the average, the solvent increases the magnitude of induced dipoles by -13% (mean of the distribution), and changes the orientation of the induced dipole component by an average of 45 degrees (Fig. 6b), with time dependent fluctuations of 5% for the ratio E_{p+w}^μ/E_p^μ , and of 25 degrees for the angle $\bar{E}_{p+w}^\mu, \bar{E}_p^\mu$ for which the distributions are not shown.

A better measure of the influence of electronic polarizability can be obtained from the analysis of its effect on the *total* local field. Distributions of the time averaged ratios $\langle E^t/E^q \rangle$ where E^t is the magnitude of the local *total* electrostatic field (containing contributions from both permanent and induced dipoles), and E^q the magnitude of the local field due to permanent dipoles alone (see also Eqn 3) have therefore been analyzed. These distributions are not shown here since they closely resemble those previously obtained (21). They confirm that the average contribution from electronic polarizability to the local *total* fields inside the protein is small on the average, with the means of the distributions ranging between 7-14%, but due to a broad tail of the distribution towards higher values, it remains significant for an appreciable fraction of the atoms.

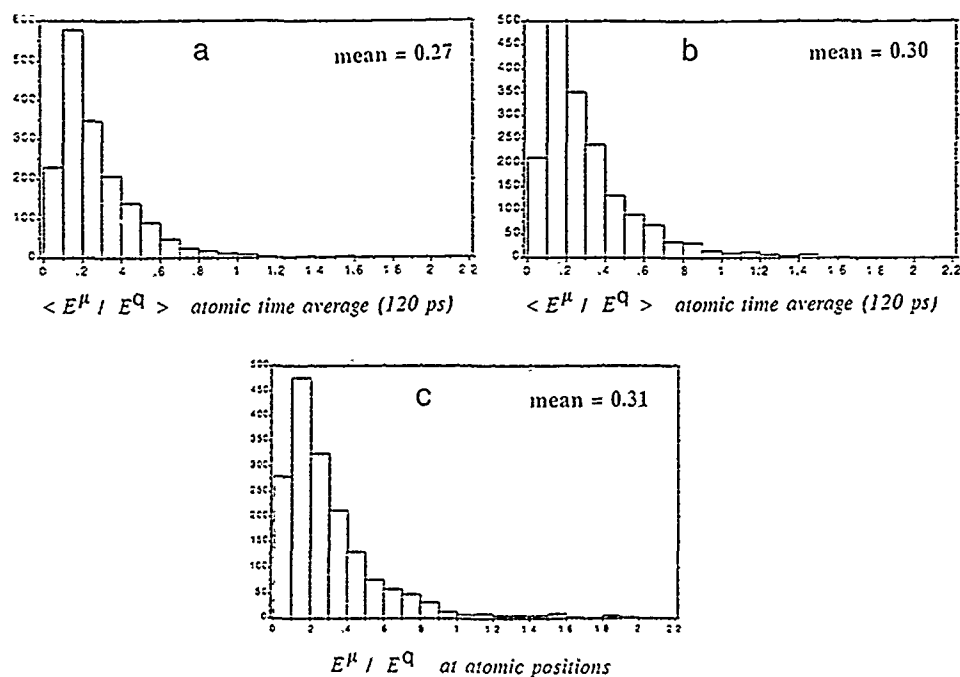


Fig 4: Statistical analysis of the relative magnitudes of local electrostatic fields due to induced dipoles \bar{E}^μ , and to permanent dipoles (or partial charges) \bar{E}^q , computed at atomic positions in barnase.

(a) Distribution of the ratios between magnitudes of fields due to induced dipoles, and fields due to partial charges, computed at atomic positions i and averaged over the 120 ps trajectory generated by the simulation. Both \bar{E}^μ and \bar{E}^q include contributions from surrounding solvent. However, \bar{E}^μ does not include contributions from *induced* water dipoles.

(b) Same as in (a) but, \bar{E}^μ and \bar{E}^q are computed from the barnase trajectory without including contributions from surrounding water molecules.

(c) Distribution of the ratio E^μ/E^q in the static crystal structure of barnase in vacuum. All calculations are performed with $\epsilon=1$, full charges on ionizable groups and a cutoff distance of 8 Å together with the shifting function described in the text.

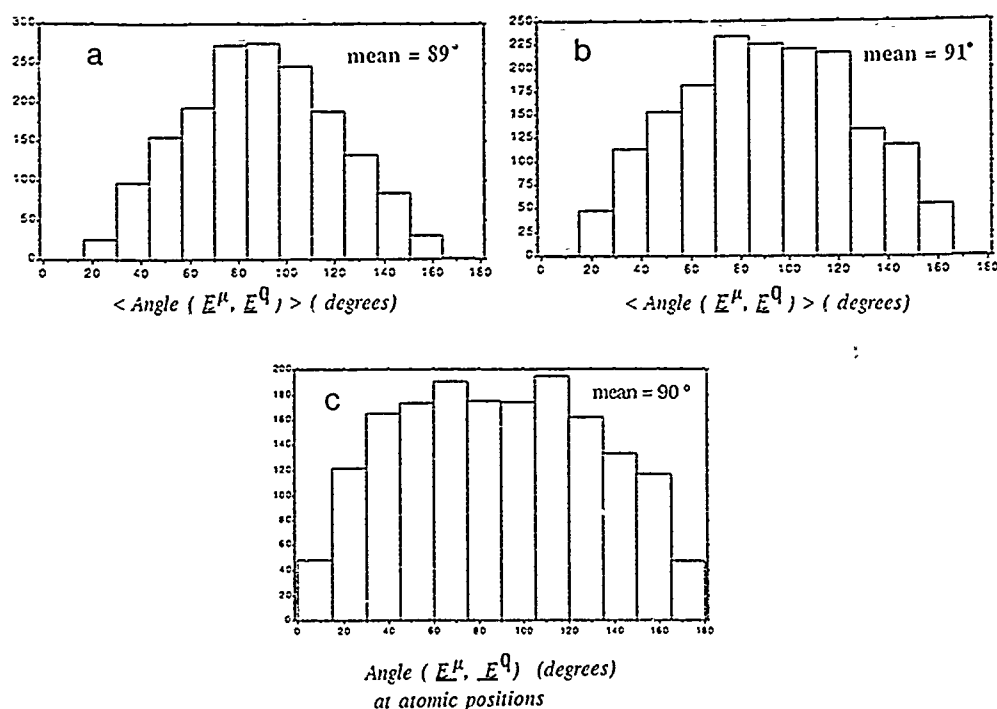


Fig 5: Relative orientations between local field components due respectively to induced dipoles \tilde{E}^μ and permanent dipoles \tilde{E}^q in barnase.

(a) Distribution of the average angle between the field vectors computed from the last 25 ps of the simulation in presence of water molecules. (b) As in (a) but water contribution to electrostatic fields is not included. (c) Distribution of the angles between the same field components computed in the static crystal structure of barnase in vacuum with computational parameters as for the results presented in Fig. 4c.

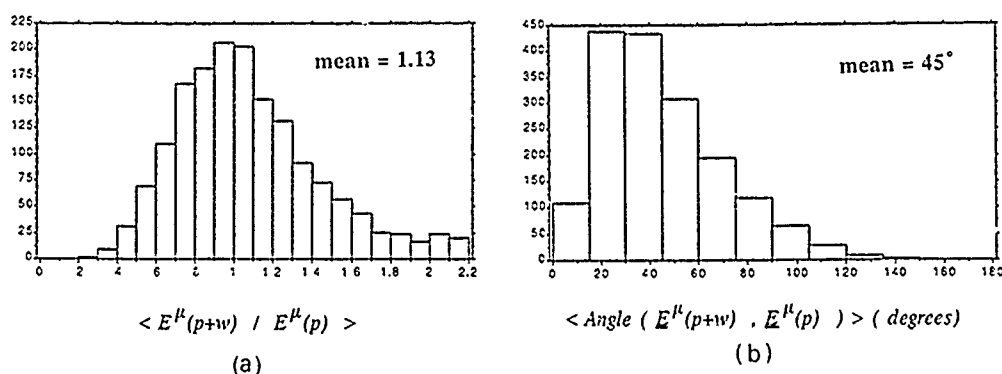


Fig 6: Influence of solvent on the induced dipoles inside the protein. (a) Distribution of the average ratio of induced field magnitudes at atomic positions computed along the 120 ps trajectory of barnase: $\langle E^\mu_{p+w} / E^\mu_p \rangle$, where the numerator is the induced field on an atom i in presence of all protein and solvent atoms, while the denominator is the same field computed in absence of solvent. (b) Distribution of the average angle between the corresponding induced field vectors described in (a).

Contributions from solvent polarization to local fields inside the protein

A correct description of the dielectric properties and electrostatic interactions in solvated proteins requires proper evaluation of contributions from surrounding solvent molecules. These contributions have as origin the highly polar character of the water molecules which will both reorient in response to the field generated by the protein and electronically polarize their surroundings. In addition to these, of course, water also exerts an indirect influence through its effects on protein structure and dynamics. Our MD simulations performed with a detailed microscopic representation of both the protein and the water portions, together with the procedure of computing induction effects, are ideally suited for obtaining a detailed picture of these effects. Data on the influence of water molecules on induced dipoles in the protein was presented in the previous section. Here, we turn our attention to the contribution of surrounding water to electrostatic fields due to partial charges inside the protein. To evaluate this contribution we compute time averages of the ratios of electrostatic field magnitudes

$\langle E_{p+w}^q/E_p^q \rangle$ at atomic positions in the protein for conformations along the generated

120 ps trajectory. E_{p+w}^q and E_p^q are respectively, the local field magnitude generated by permanent dipoles of the protein and the surrounding solvent and the local field magnitude generated by protein dipoles alone. Note that effects of electronic polarizability are not included here. The distribution of these ratios is shown in Fig. 7a and the corresponding standard deviations (or fluctuations) are shown in Fig. 7b. From the distribution in Fig. 7a we see that water increases the magnitude of local fields at protein atoms by a substantial amount: 28% on the average (mean value of 1.28 for the averaged ratios of field magnitudes) with a sizable fraction of cases displaying increases of 100% or higher! The standard deviations of the same field ratios are also large with a mean value of 0.44 (Fig. 7b). Surrounding water is also found to have an appreciable effect on the orientation of local fields due to permanent dipoles inside the protein. This is illustrated by the distribution of the average relative orientation between local field vectors due respectively to permanent dipoles of the entire system (protein + water) and to those of the protein alone (Fig. 7c). This distribution has a mean of 32 degrees, but it is broad with a substantial tail reaching values as high as 140 degrees, indicating that surrounding water can in some cases nearly reverse the direction of the local field. These results confirm the very important contribution of surrounding water to electrostatic fields inside the protein, and that the major part of this contribution is, as expected, in the component due to permanent water dipoles. The broad shapes of the computed distributions and preliminary attempts to relate them to structural parameters lead us to assume that solvent contributions would not be evenly distributed throughout the protein matrix. Structural features of the protein, degree of exposure to solvent as well as structural and/or dynamic properties of surrounding water molecules may strongly influence these contri-

butions. Indeed, preliminary analysis shows (see Table 2) that in exposed parts of the protein (atoms that expose more than 90% of their surface to solvent) the contribution from water polarization to the magnitude and orientation of local fields reaches respectively 60% and 43 degrees on the average. However, in buried regions (atoms that expose less than 10% of their surface to solvent) this contribution is reduced on the average to 12% (in field magnitude) and 20 degrees (in field orientation). On the other hand, the contribution from electronic polarizability which averages 13% in field magnitude and 12 degrees in field orientation, is found to be insensitive to the degree of exposure to solvent. Hence, electronic and solvent polarization effects are (on the average) of comparable size and importance in buried regions, while in exposed regions solvent polarization, as expected, clearly dominates.

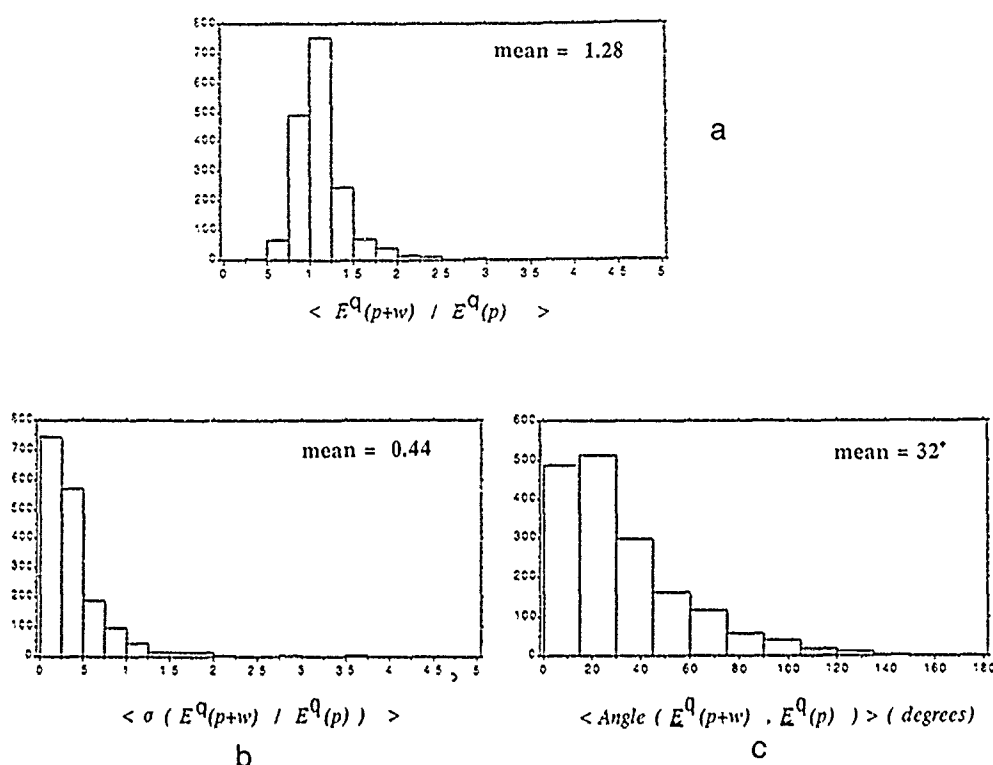


Fig 7: Contributions of water molecules to local fields due to permanent dipoles and charges in barnase. (a) Distribution of the average ratios between field magnitudes E^Q_{p+w} , the local field due to the permanent dipoles of protein and water, and E^Q_p the local field due to permanent dipoles of the protein alone. Fields are computed at atomic positions and averaged along the entire 120 ps trajectory of the simulation. (b) Distribution of the fluctuations of the ratios described above about the average values. (c) Distribution of the average angles between the local field vectors described in (a).

TABLE 2

Electrostatic interactions in barnase: Contribution from induced dipoles and solvent

		Flexible protein in vacuum	Flexible protein in solvent	Rigid crystal protein
Angle \vec{E}^l, \vec{E}^q	(a)	12.1 (6.4)	9.1 (5.4)	12.9(14)
	(b)	13.1 (6.6)	12.4 (6.5)	
	all	12.7 (6.7)	11.5 (6.5)	
Angle \vec{E}^μ, \vec{E}^q	(a)	91.9 (24.1)	91.2 (27.6)	89 (43)
	(b)	87.9 (23.7)	84.3 (24.9)	
	all	90.8 (23.6)	89.1 (26.0)	
Angle $\vec{E}_{p+w}^q, \vec{E}_p^q$	(a)		42.7 (18.4)	
	(b)		20.2 (10.7)	
	all		32.1 (15.2)	
Angle $\vec{E}_{p+w}^\mu, \vec{E}_p^\mu$	(a)		54.7 (25.3)	
	(b)		32.8 (17.8)	
	all		45.2 (21.8)	
E^l/E^q	(a)	1.12 (0.40)	1.12 (0.44)	1.07 (0.30)
	(b)	1.12 (0.38)	1.16 (0.41)	
	all	1.13 (0.40)	1.14 (0.45)	
E^μ/E^q	(a)	0.28 (0.13)	0.22 (0.12)	0.31 (0.30)
	(b)	0.31 (0.15)	0.30 (0.15)	
	all	0.30 (0.14)	0.27 (0.14)	
E_{p+w}^q/E_p^q	(a)		1.56 (0.63)	
	(b)		1.12 (0.26)	
	all		1.28 (0.44)	
E_{p+w}^μ/E_p^μ	(a)		1.15 (0.51)	
	(b)		1.08 (0.34)	
	all		1.13 (0.50)	

120 ps MD simulations of Barnase, July 1989

The table summarizes the results obtained. It lists mean values and fluctuations (in parentheses) of computed atomic averages of quantities for which the distributions are shown in Fig. 4-7.

The first column gives the expression of the computed quantity using the same conventions as in the text and in the Figures. Columns 2 and 3 list results obtained from distributions computed with barnase trajectories, respectively in absence and presence

of water molecules. Column 4 lists results obtained from computations performed on the static crystal structure of barnase in vacuum. (all) designates values obtained using all atoms in the protein. (a) designates values obtained for atoms that expose more than 90% of their surface to solvent (a total of 392 atoms), and (b) values obtained for buried atoms (446 in number) – with less than 10% of their surface accessible to solvent.

Note that the computed standard deviations (given in parentheses in column 4) concern spatial fluctuations and not time dependent fluctuations as in columns 2 and 3 and are therefore of a different nature.

CONCLUDING REMARKS

In this account, recent results have been presented on 120 ps Molecular Dynamics simulations of a small protein barnase in presence of explicit water molecules. Although detailed analysis of the trajectories is still in progress, evidence has been presented that it provides a reasonable description of the protein and water portions. The deviations of the average protein conformation from the starting crystal structure is reasonably low (~ 1.1 Å rms for backbone atoms), and the agreement between computed and crystallographic atomic fluctuations is satisfactory for portions of the protein that do not participate in crystal contacts. The structure of the water around polar and non-polar groups on the protein surface also seems reasonable in that it agrees well with previous observations made either in hydrated crystals or in other simulations of small solvated molecules.

The thrust of our study concerned the use of the generated trajectories to obtain a detailed microscopic description of different contributions, including electronic polarizability, and solvent effects to electrostatic interactions in a solvated protein. Table 2 summarizes the highlights of the results presented here. A number of clear trends emerge. The contributions of surrounding water to electrostatic fields due to permanent dipoles inside the protein is substantial. It affects field magnitudes and field orientations respectively, by 28% and 32 degrees on the average. In comparison, the contribution from electronic polarizability alone is much lower with an average of -10% in field magnitude, and -12 degrees in field orientation. However, analysis of how these effects are distributed, shows that they display an appreciable degree of inhomogeneity throughout the protein matrix. This is confirmed by the finding that solvent polarization effects are on the average twice the sizes of induced dipole effects in exposed regions, while in buried regions the two contributions are (on the average) of comparable size. Other factors such as the precise local constellation of polar and non polar groups should further modulate the relative importance of electronic polarizability versus water polarization. Further analysis is in progress to illustrate this point.

Our study represents substantial progress towards obtaining a detailed microscopic description of electrostatic interactions in proteins. But it is only a first step in an on-going effort, since many problems still need to be overcome. Simulation time

scales need to be extended, possibly to the nanosecond range, several parallel simulations should be performed to reduce bias from initial conditions, and induced polarizability must be included in the Molecular Dynamics simulations, not to mention the need to rederive force-fields and energy parameters which are physically consistent with a detailed microscopic description.

As computational power is increasing exponentially, our ability to tackle these problems is in constant progress. We hope to be able to produce better trajectories of solvated proteins in the near future, and to extract from them improved descriptions of electrostatic properties which we intend to confront with those obtained by other methods such as the Langevin dipole treatment for water (8) and various macroscopic continuum approaches (13) (17).

REFERENCES

- 1 A. Warshel, *Biochemistry*, **20** (1981) 3167
- 2 J.M. Baldwin, *Prog. Biophys. Mol. Biol.*, **29** (1975) 225
- 3 P.K. Weiner, R. Langridge, J.M. Blany, R. Schaefer R. and P.A. Kollman, *Proc. Natl. Acad. Sci. USA*, **79** (1982) 3754
- 4 J.A. Tainer J.A., E.D. Getzoff, J.S. Richardson and D.C. Richardson, *Nature (London)* **306** (1983) 284
- 5 M.F. Perutz, *Science*, **201** (1978) 1187
- 6 J.M. Thornton, *Nature*, **295** (1982) 13
- 7 B. Gelin and M. Karplus, *Proc. Natl. Acad. Sci. USA*, **72** (1975) 2002
- 8 A. Warshel and M. Levitt, *J. Mol. Biol.*, **103** (1976) 227
- 9 P. Ahlstrom, O. Teleman, B. Jonsson, *J. Amer. Chem. Soc.* **110** (1988) 4198
- 10 C.L. Brooks and M. Karplus, *J. Mol. Biol.* **208** (1989) 159
- 11 M. Levitt and R. Sharon R., *Proc. Natl. Acad. Sci. USA*, **85** (1988) 7557
- 12 C. Tanford and J.G. Kirkwood, *J. Amer. Chem. Soc.*, **79** (1957) 5333
- 13 J. Warwicker and H.C. Watson, *J. Mol. Biol.*, **157** (1982) 671
- 14 R.J. Zauhar and R.S. Morgan, *J. Comput. Chem.*, **9** (1988) 171
- 15 M.K. Gilson, K.A. Sharp and B.H. Honig, *J. Comp. Chem.*, **9** (1987) 327
- 16 M.K. Gilson, A. Rashin, R. Fine and B.H. Honig, *J. Mol. Biol.*, **183** (1985) 505
- 17 I. Klapper, R. Hagstrom, R. Fine, K. Sharp and B.H. Honig, *Proteins*, **1** (1986) 47
- 18 M.K. Gilson and B.H. Honig, *Proteins*, **3** (1988) 32
- 19 M.J.E. Sternberg, F.R.F. Heys, A.J. Russel, P.G. Thomas and A.R. Fersht, *Nature*, **330** (1987) 86
- 20 N.K. Rogers and M.J.E. Sternberg, *J. Mol. Biol.*, **174** (1984) 527
- 21 D. Van Belle, I. Couplet, M. Prévost and S.J. Wodak, *J. Mol. Biol.*, **198** (1987) 721
- 22 S.T. Russel and A. Warshel, *J. Mol. Biol.*, **185** (1985) 389
- 23 T.P. Lebrand and P.A. Kollman, *J. Chem. Phys.*, **83** (1986) 2923
- 24 P. Delhaise, D. Van Belle, M. Bardiaux and S.J. Wodak, *J. Mol. Graphics* **3** (1985) 116
- 25 W.L. Jorgensen, *J. Am. Chem. Soc.*, **103** (1981) 335
- 26 Y. Maugen, R.W. Hartley, E.J. Dodson, G.G. Dodson, G. Bricogne, C. Chothia and A. Jack, *Nature*, **297** (1982) 162
- 27 P. Alard, *Mémoire de Licence, Université Libre de Bruxelles* (1982)
- 28 C.L. Brooks, B.M. Pettitt and M. Karplus, *J. Chem. Phys.*, **83** (1985) 5897
- 29 M. Prévost, G. Lippens, D. Van Belle and S.J. Wodak, *J. Mol. Phys.* (1989) (submitted)

- 30. P.P. Ewald, Ann. Physik, **64** (1921) 253
- 31. Mc Lachlan A.D., J. Mol. Biol., **128** (1979) 49
- 32. See f.e in G.H. Stout and L.H. Jensen, X-ray structure determination, Macmillan Company, New-York (1968) 206
- 33. S.J. Wodak, P. Delhaise, P. Alard and C. Renneboog-Squilbin, J. Mol. Biol. **181** (1985) 317
- 34. B.K. Lee and F.M. Richards, J. Mol. Biol., **55** (1971) 379
- 35. P. Rossky and M. Karplus, J. Am. Chem. Soc., **101** (1978) 1913
- 36. R. Taylor, O. Kennard and W.J. Versichel, J. Am. Chem. Soc. **106** (1984) 224
- 37. S. Swaminathan, S.W. Harrison and D.C. Beveridge, J. Am. Chem. Soc., **100** (1978) 5705
- 38. W.L. Jorgensen, J. Chem. Phys., **77** (1982) 5787
- 39. B.R. Brooks, R.E. Bruccoleri, D. Olafson, D. States, S. Swaminathan and M. Karplus, J. Comp. Chem., **4** (1983) 187
- 40. L. Verlet L., Phys. Rev., **159** (1967) 98
- 41. J.P. Ryckaert, G. Ciccotti G. and H.J.C. Berendsen, J. Comp. Phys., **23** (1977) 327

DISCUSSION

SURCOUF - When you performed your calculations on the crystal structure, did you take into account the water molecules of the crystal structure ?

WODAK - The Molecular Dynamics simulations included 94 water molecules whose positions were determined crystallographically as well as 2265 additional water molecules whose starting positions were generated on a cubic lattice dimensioned to yield unit density.

Calculations performed on the static crystal structure of barnase did not include any water molecules, as the purpose of these calculations has been to analyze the intrinsic properties of the protein matrix.

ANGYAN - I should like to comment the implementation of the induced dipoles model in MD simulation. The most difficult problem is the calculation of forces. One has essentially two probabilities that works :

- a) do not iterate at all : then the forces do not include the effect of field gradient of the induced moments on induced moments.
- b) iterate up to self-consistency : then the above term appears in the force expression. Nevertheless one should be extremely careful in the third case, in the not-fully-iterated case, because there is a risk of non-conservation of the energy. The reason is that for not self-consistent induced moment the "obvious" force formula is not the exact derivative of the energy expression.

WODAK - Indeed, using an iterative self-consistent procedure to compute induced dipoles and fields is equivalent to numerical computations of energy derivatives as opposed to exact analytic procedures, and can therefore cause energy conservation problems. Very stringent convergence criteria must be used to prevent such problems. Better ways to incorporate self-consistent effects which do not suffer from these shortcomings require modifying the equations of motions in ways that do not significantly affect the actual dynamics of the system.

PEPE - Are the electrostatic effects all related to a choice of charge distribution ?

WODAK - Yes, they are indeed. We have shown in our published study (Van Belle et al, 1987) that the use of different values of electrostatic charges on ionizable groups in proteins influences the size and directions of induced dipoles to an extent which is comparable or larger in magnitude than the induction effect itself. It may be expected that the influence from changing the values of permanent dipoles and charges should be reduced in presence of solvent due to screening effects, but the relevant computations need still to be done.

SOUIMPASIS - In the MD calculations you use a cutoff of 7 Å (for computational convenience) and compare the result with the static calculation where there is no cutoff, why don't you use the same cutoff in both calculations ?

WODAK - The published work on the analysis of induction effects in proteins reported results obtained using an infinite cutoff distance for including pairwise interactions. But, in the present account, where results from calculations performed on computer generated trajectory, and on the static crystal structure are compared, the same cutoff distance and switching/shifting potential for long-range interactions have been used.

BUCKINGHAM - You indicated that the electric field arising from water molecules tend to be perpendicular to the field of the permanent dipoles of the protein. This contrasts with the Onsager reaction field model in which the field due to the polarized environment is parallel to the primary dipole.

WODAK - There must be a misunderstanding. I indicated that the local field components arising from induced electronic polarizability, and not the fields arising from the presence of water molecules, tend to be perpendicular to the local fields generated by the permanent dipoles of the system.

KWIATKOWSKI - Several scientific groups try to construct atom-atom potentials (in fact, these potentials depend on the choice of several parameters). You have used in your

calculation the parameters from CHARMM. To what extent your quantitative predictions would be changed if you would use another potential ?

WODAK - The molecular trajectories generated by Molecular Dynamics will obviously depend on the energy functions and values of energy parameters. The results of the detailed electrostatic calculations will more particularly depend on values of partial charges and on values of atomic polarizabilities (as shown in our published work - Van Belle et al., 1988, the dependence on the former is more dramatic). We believe however, that the general conclusions reached here, namely, (1) that contributions to electrostatic fields from surrounding solvent, are on the average three times the size of contributions from electronic polarizability and (2) that the relative contributions from solvent and electronic polarizability are highly dependent on the local environment, would hold, when we use other force-fields representative of protein structures.

DURUP - I could not see whether your pictorial description of water molecules allows to characterize the features of the first and second water layers, bulk water, water clusters, etc.

WODAK - The pictorial representation of water structure around specific protein groups (side-chains) allows a visual analysis mainly of the first layer of water molecules, and that only in a relatively limited local region.

PERAHIA - Did you analyze the variation of your statistics when you use trajectories started with different initial conditions ? I ask this question because the electrostatic energy is sensitive to ample motions, and most of them are not necessarily present in a single trajectory of 120 ps.

WODAK - Being severely limited by computer time we did not have the possibility so far to generate more than just one trajectory, although we are much aware that data from multiple trajectories should be very useful.

SOUMPASIS - What is the best effective dielectric constant you would suggest for a protein in water on the basis of your MD work ?

WODAK - This question cannot be readily answered at present. Our calculations have shown that neither $\epsilon=3$ nor $\epsilon=r_{ij}$ (where r_{ij} is the distance between a pairs of atoms i and j) can properly reproduce electrostatic energies computed using a detailed atomic representation (using $\epsilon=1$ and including polarizability effects). A similar analysis has not as yet been performed in presence of surrounding water. Most importantly, our analysis demonstrates that the protein matrix is rather heterogeneous electrostatically and argues strongly against the use of the same effective dielectric constant throughout.

ELECTROSTATIC FREE ENERGY AS THE FUNDAMENTAL STRUCTURE FUNCTION CORRELATOR IN PROTEINS : SOME PERSPECTIVES FROM MICROSCOPIC SIMULATIONS OF PROTEINS FUNCTIONS

A. WARSHHEL

Department of Chemistry, University of Southern California, Los Angeles,
California, 90089-1062 (U.S.A.)

INTRODUCTION

Establishing a clear correlation between protein structure and function is one of the major goals in experimental as well as theoretical molecular biology. It is becoming widely recognized that the concept of electrostatic free energy plays a key role in translating a particular structural pattern to specific functional properties (1-4). The efforts from our laboratory in the past 15 years have to a large extent been invested in a microscopic verification of this point of view (see for example, ref. 1, 5, 6-11). These studies, which will be briefly reviewed in what follows, involve the development of a simplified microscopic model [which provided the first semi-quantitative way for studying the energetics of solvated proteins (1,5)] as well as more advanced approaches that have become practical upon the emergence of modern powerful computers. The implementation of these strategies in studies of several different types of protein functions is reviewed in the subsequent sections.

EVALUATION OF ELECTROSTATIC FREE ENERGIES IN SOLVATED PROTEINS

In order to understand biological processes on a molecular level it is crucial to somehow be able to estimate relevant electrostatic free energies. This could not readily be accomplished by macroscopic models since these inevitably must rely on phenomenological dielectric constants (whose relation to the actual microenvironment is not given by macroscopic considerations). The point of view taken in this laboratory quite early, was to avoid the macroscopic formulation altogether and instead attempt to follow a route based on microscopic

considerations. The enormous dimensionality of electrostatic problems in proteins and the long-range nature of these effects have motivated the development of simplified models (1,5), which still capture the major contributions to electrostatic energies. The Protein Dipoles Langevin Dipoles (PDL) model is an example of this type of approach, and it is based on the working hypothesis that the electrostatic energy of the average structure of the system is a good approximation for the electrostatic free energy of the system. The average structure of the protein is usually taken as the experimentally determined conformation from X-ray crystallography, and the contributions from groups of the protein to the electrostatic energy are evaluated on a microscopic level where permanent residual charges and induced dipoles are taken into account. The solvent around the protein is treated by considering its average polarization as given by the Langevin dipole model and representing the solvent molecules by a grid of Langevin type dipoles [see (12) for details].

The philosophy behind this simplified water model is as follows. The average polarization of any given water molecule near an ion is related in some way to the field from the ion. If we knew the distribution function for this average polarization we could evaluate the ion solvation free energy by summing the scalar products of the average dipole at each grid point and the field from the ion at this point, and then adding the free energy of polarizing the solvent dipoles (this polarization costs about half of what is gained from the field-dipoles interactions). Fortunately one can determine the polarization of water molecules as a function of the applied field from microscopic simulations (1c) (and/or by refining the model by fitting calculated and observed solvation energies (5)) and represent the resulting polarization function by a scaled Langevin-type of expression. By surrounding the solute (or protein) with a grid of polarizable dipoles obeying the expected solvent polarization, it appears that free energies associated with solvation can be reasonably well produced without the need for an all atom water model. The PDL scheme, which treats both the solvent and the protein polarization explicitly, thus provides a simple and effective way of assessing electrostatic energies in solvated proteins.

With the recent advances in computing power it has become possible to sample the phase space of highly multidimensional systems by direct simulation approaches such as MC and MD methods. Consequently, one can explore formal statistical mechanical calculations, which earlier had been beyond reach, but now can be employed (with serious convergence problems) even for systems of considerable size (9,10,13-16,26,36,38). With these so called "free energy perturbation" (FEP) methods (17a,b) one can directly evaluate free energy differences from MC or MD trajectory calculations. This is usually done by introducing a mapping parameter (controlling the effective potential surface in the simulation) with which the system can be gradually driven from a particular

reference state to a state whose free energy one wishes to calculate (such an approach was first introduced for ion solvation problems in ref. 13a and for enzymes in ref. 13b). In this type of calculation it has become practically possible to treat a large part of the system microscopically, by explicit all-atom representations. An inherent difficulty, however, in performing such microscopic simulations on large molecular systems is associated with the selection of appropriate boundary conditions. In this respect, the use of periodic boundary conditions in combination with simple truncation of interactions beyond a certain range or in combination with Ewald summation has become the most popular prescription. However, it is clear that in some cases the introduction of artificial periodicity or symmetry in the calculations can lead to rather unphysical representations of the relevant electrostatic fields [for an excellent discussion of these problems, see ref. (17c)].

Considering the above mentioned problems we have concentrated on developing non periodic models. In the first attempt to evaluate solvation free energies in a charge transfer reaction by an all-atom solvent model using FEP/MD simulations (13), the dimensionality of the problem was reduced by applying surface constraints [introduced originally in the SCSSD model (18)] to the solvent molecules. This was done by using a limited number of water molecules surrounded by a spherical surface layer of molecules that represents the effect of the surrounding in a corresponding infinite system. A more refined model incorporating explicit dynamical polarization constraints was subsequently reported (19), referred to as the Surface Constraint All Atom Solvent (SCAAS) model. This model uses a limited number of waters in order to create a sphere of solvent around the relevant groups (the radius of this sphere is typically 10-15 Å). Angular and radial constraints are incorporated in order to compensate for the artificial surface created as a result of using a finite number of water molecules. The extension of the SCAAS model to free energy calculations of solvated proteins [and its use as a substitute for the earlier PDL (1,5) and SCSSD (18) models] is quite straightforward (20). The SCASS model has been used extensively in this laboratory in recent years and appears to provide an efficient way to carry out microscopic simulations of solvated proteins and chemical processes in solution. However, despite its ability to overcome some of the problems associated with imposing artificial symmetry on the system, the method still requires the truncation of long-range interactions in order to reduce computational costs. Unfortunately, when treating larger systems it is of great importance to use a large cutoff radius for water-water interactions in order to account for long-range electrostatic correlations. A step towards this dilemma is the recently developed extension of the Ewald method to non-periodic systems (21), which may allow for the practical implementation of SCAAS systems of much larger size than the ones currently used. At any rate, we presently find (by

using calculations of intrinsic pKa's as a general bench mark) that the accuracy of FEP calculations is still not much better than that of the PDL method (in some cases the accuracy is less than that of the PDL model). This might reflect the fact that present simulation times allow us to sample an extremely small part of the relevant phase space and of course the above mentioned long range problem.

SIMULATIONS OF ENZYME CATALYSIS

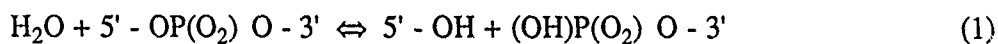
Our studies of enzymatic reactions have been largely based on the notion that the *difference* between enzyme and solution reactions is mainly determined by the corresponding surroundings of the reacting systems and not so much by the actual quantum mechanics of the "solute" atoms. Apparently, purely quantum mechanical approaches to enzyme catalysis, that do not incorporate the environmental effects from the protein into the solute hamiltonian, may lead to rather irrelevant results. These considerations led to development of SCF-MO approaches (5) (MINDO-2 and QCFF/ALL) which combined the microscopic potential of the protein and its surrounding solvent into the quantum mechanical hamiltonian. In later stages of these studies (6,13) it was concluded that the Valence Bond (VB) method had better chances of producing conclusive results (since it can easily be calibrated with solution experiments in a much more unique way than other methods). This led to the development of the Empirical Valence Bond (EVB) method which we will be only briefly review here.

The introduction of the EVB scheme for studies of enzymatic reactions was based on the observation that many chemical reactions can be described in terms of elementary "resonance structures" (RS) that specify the charges and bonding of the reacting atoms at different VB configurations ; e.g. a heterolytic bond cleavage is described by (X-Y) and (X⁻ Y⁺). Furthermore, the different RS can be conveniently represented by standard analytical force fields possibly augmented by gas-phase quantum mechanical calculations. With this type of description one can explore chemical reactions in enzymes by evaluating the energetics of the relevant RS in solution and in the protein's active site (the EVB method is easily implemented within the PDL as well as the FEP/MD framework). For instance, a proton transfer reaction in an enzyme can be examined by considering the energetics of the two RS (A-H B) and (A⁻ H-B⁺). If the protein is designed to catalyse such a reaction it will stabilize the (A⁻ H-B⁺) state relative to (A-H B) more than water does. Similar treatments can be introduced to study more complicated enzymatic reactions (see ref. 10 for a detailed description).

The EVB approach has been applied to several important families of enzymes (1,6,8-10,26). In the cases where comparisons were made between the PDL and the FEP/MD treatments, the two methods gave similar results. For

example, let us consider the interesting case of the Asp 32 \rightarrow Ala mutation in subtilisin (11). This mutation which removes the catalytic Asp 32 leads to a four orders of magnitude reduction in the reaction rate (22). The calculations reported in ref.11 on this mutation are summarized in ref. 11. The main effect of the catalytic is reproduced by both the PDL and FEP/MD methods. Both calculations indicate that Asp 32 contributes to catalysis by electrostatic stabilization of the transition state and not by the double proton transfer mechanism depicted in many textbooks. The calculations above as well as related studies on other enzymes lend support to the proposal that enzymes catalyse their reactions by providing active site environments which are electrostatically complementary to the *change* in charge distribution of the reacting system associated with forming the relevant transition states (1,6,10,11).

The importance of electrostatic energy is particularly dramatized in metal catalysed reactions, where the enzyme often reduces activation barriers by more than 10 kcal/mol. An illustrative example of this is provided by the enzyme staphylococcal nuclease (SNase) which catalyses the hydrolysis of both DNA and RNA at the 5' position of the phosphodiester bond (23-25) :



EVB studies of this system (26) reproduce the overall observed catalytic effect of the enzyme and demonstrate the crucial role of the metal charge in stabilizing the hydroxide ion and reducing the energetics of proton transfer from water to the enzyme general base. The calculations also reproduce and explain the effect of metal substitution on the corresponding catalytic activity. This indicates that the EVB method captures the first order effects in enzyme catalysis by focusing on the relevant environmental effects (the change in electrostatic energy upon moving from water to the enzyme active site).

ION CHANNELS

The energetics of ion transport through membrane channels is a rather complex problem since the interacting system comprises several different phases. In addition to the membrane spanning protein(s), with internal solvent molecules, both the surrounding aqueous and lipid phases contribute to the overall solvation of a migrating ion. The most widely studied system for channel mediated ion transport is the Gramicidin A (GA) channel. Because the activation energies for ion permeation (5-7 kcal/mol) are known from direct experimental data (29,30), the difficulties associated with calculations of energy profiles along the channel immediately becomes apparent. Although continuum calculations that assumed a large dielectric constant inside the channel could be adjusted to reproduce the observed barrier (31,32), the results of extensive microscopic simulation studies

have been rather disappointing, at least as far as activation barriers are concerned. This appears mainly to be due to the neglect of some of the main electrostatic contributions, while treating other factors more or less rigorously (33,34). In particular, it appears that models which neglect either the bulk water phase or the lipid membrane cannot account for the fairly high permeation rates that are observed experimentally (for an overview of early calculations on the GA system, see ref. 35).

We have recently reported free energy calculations on the Na^+ -GA-water-membrane system using both the PDL and the FEP/MD methods (36), which appear to resolve the problems outlined above. In these calculations the membrane surrounding the GA channel was represented by a large grid of polarizable point dipoles (with no permanent dipole moments), giving a dielectric constant of $\epsilon_m = 2$ for the membrane. By using the PDL method in combination with energy minimization (EM) a complete energy profile for the permeation of a Na^+ ion was calculated.

It has been found that the calculated values for the energy barriers agree fairly well with those observed experimentally. Furthermore, the calculations indicate that the contribution to the ion solvation energy in the channel from the polarization of the surrounding membrane and (to a lesser extent) from the induced dipoles of the GA helix can amount to as much as ~ 10 kcal/mol.

While FEP/MD simulations for this system would be very expensive for calculating a complete free energy profile (in particular at the channel entrances where a large number of water molecules would have to be included), it is of interest to perform such a calculation in the interior of the channel. In this region only a limited number of waters are needed to account even for rather long-range solvation effects and the results can be directly compared to those obtained with the PDL/EM scheme in the centre of the channel. The results of such calculations (36) produced a difference of 2.3 kcal/mol between the free energy of the ion in the centre of the channel and in water. This value is similar to the corresponding result from the PDL free energy is -2 to +3 kcal/mol (36).

The study of ion solvation in the GA channel clearly illustrates the danger in neglecting certain parts of the system for which electrostatic energies are evaluated. In this case, discarding the induced membrane polarization appears to be equivalent to introducing an error of ~ 10 kcal/mol and it is therefore not surprising that most reported energy profiles have given too high energies in the interior of the channel.

ELECTROSTATIC CONTROL OF THE CHARGE SEPARATION PROCESS IN PHOTOSYNTHETIC SYSTEMS

The recent determination of the X-ray structure of bacterial reaction centres (37) provides an opportunity of probing the detailed molecular nature of

photosynthetic processes. In particular, one can now attempt to explain the relation between the protein structure and the energetics of the primary charge separation event. A major challenge in this task is the puzzling directionality of the electron transfer process in bacterial reaction centers. That is, the X-ray structure reveals a bacteriochlorophyll dimer P, which serves as the primary electron donor, two neighboring accessory bacteriochlorophylls (B_L and B_M) and two bacteriopheophytins (H_L and H_M). These chromophores are arranged in a seemingly very symmetric manner in two subunits (L and M) forming the $(PBH)_L$ and $(PBH)_M$ branches. Despite this nearly symmetric arrangement it appears that the primary charge separation process occurs almost exclusively via the L branch, while the electron transfer pathway through the M branch is virtually blocked. This finding offers an exciting opportunity of understanding the control of electron transfer pathways by protein microenvironments. Furthermore, this problem is a particularly useful test case for various methods for correlating structure and function in proteins.

With the X-ray structure of the reaction centre we can calculate the free energy of the various possible charge transfer states along the charge separation coordinate. This can be done by either the PDL or the FEP methods provided, however, that the calculations include all the key electrostatic elements (protein permanent dipoles, protein induced dipoles and surrounding water and membrane). Furthermore, meaningful calculations require that one takes into account the energy of forming the relevant charged chromophores in the gas phase. This energy (which is not given in a reliable way by current quantum mechanical methods) can be obtained by using experimentally observed redox and oxidation potentials and the energies of the different charge transfer states in solution (38). We thus use a thermodynamic cycle and obtain the relationship

$$\Delta G_i^P = \Delta G_i^W + \Delta \Delta G_{i, \text{sol}}^{W \rightarrow P} \quad (2)$$

where ΔG_i^P is the free energy of forming the indicated state in its protein site, while $\Delta G_{i, \text{sol}}^{W \rightarrow P}$ is the change in the "solvation" free energy of the given state upon moving the chromophores from water to their protein sites.

The PDL calculations of ΔG_i^P for the two branches show a clear preference for the L branch with a significant activation barrier for charge separation through the M branch (39). These findings indicate that the protein microenvironment provides more electrostatic stabilization to the $(P^+B^-)_L$ state than to the $(P^+B^-)_M$ state, thus contributing to the control of the charge separation process and facilitating electron transfer through the L branch (additional control of the electron transfer pathway may be provided through the electronic coupling matrix elements). The calculations also indicate that few amino acids (in

particular Tyr M208) are responsible for the difference between the two branches. This prediction can be examined by site specific mutations.

CONCLUDING REMARKS

In this paper we have dealt with a number of fundamental biomolecular processes in which the electrostatic free energy of the reacting species seems to play a decisive role for the overall activity. The examples given also emphasize the importance of reliable evaluation of electrostatic free energies in any attempt to examine the relationships between three-dimensional structure and function of proteins. We have indicated (and to some extent demonstrated) that calculations of energetics in proteins, in many instances, cannot become quantitative or conclusive if the treatment of electrostatic contribution is incomplete (e.g., by neglecting surrounding solvent or induced dipoles). The use of simplified representations for some contributions, within a complete treatment of electrostatic effects, appears to be more effective than using rigorous representations for only part of the system at the expense of neglecting other parts (e.g., calculating a high quality *ab initio* surface for reacting bonds while neglecting solvent around an active site, or even the active site itself).

Other studies from our lab that were not described here (10,39,40) considered the role of purely dynamical effects on the rate constants of biomolecular processes (which would manifest themselves in the magnitude of the preexponential transmission factor). It is found in most cases, with the possible exception of photobiological reactions (39,40), that such dynamical factors are of minor importance in comparison with the relevant free energy barriers. In particular, this appears to be the case for enzymatic reactions which usually involve significant activation barriers and therefore have a transmission factor that does not differ substantially from unity.

The developments reviewed here and in studies from other laboratories in this exciting field are surely not the last word in simulation of complicated biological processes. It seems clear that the enormous advances in computer technology will continue to enable changes in the simulation strategies used, and lead to yet more sophisticated and reliable methods. Nevertheless, we believe that more refined pictures of biomolecular processes that eventually will emerge will sustain the ideas about the key role of electrostatic free energies.

REFERENCES

1. (a) Warshel, A., Proc. Natl. Acad. Sci. U.S.A. 75, 5250 (1978).
(b) Warshel, A., Acc. Chem. Res. 14, 284 (1981).
(c) Warshel, A. and Russel, S.T., Q. Rev. Biophys. 17, 283 (1984).

2. Perutz, M., *Science* 201, 1187 (1978).
3. Honig, B., Hubbel, W. and Flewelling, R., *Ann. Rev. Biophys. Biophys. Chem.* 15, 163 (1986).
4. Matthew, J.B. *Ann. Rev. Biophys. Biophys. Chem.* 14, 387 (1985).
5. Warshel, A. and Levitt, M., *J. Mol. Biol.* 103, 227 (1976).
6. Warshel, A. and Weiss, R.M., *J. Am. Chem. Soc.* 102, 6218 (1980).
7. (a) Churg, A.K., Weiss, R.M., Warshel, A. and Takano, T., *J. Phys. Chem.* 87, 1683 (1983).
(b) Churg, A.K. and Warshel, A., *Biochemistry* 25, 1675 (1986).
8. Warshel, A. & Russel, S.T., *J. Am. Chem. Soc.* 108, 6569 (1986).
9. Warshel, A. & Sussman, F., *Proc. Natl. Acad. Sci. U.S.A.* 83, 3806 (1986).
10. Warshel, A., Sussman, F. and Hwang, J.K., *J. Mol. Biol.* 201, 139 (1988).
11. Warshel, A., Naray-Szabo, G., Sussman, F. and Hwang, J.K., *Biochemistry* 28, 3629 (1989).
12. Russel, S.T. and Warshel A., *J. Mol. Biol.* 185, 389 (1985).
13. (a) Warshel, A., *J. Phys. Chem.* 86, 2218 (1982).
(b) Warshel, A., *Pontif. Acad. Sci. Varia* 55, 59 (1984).
14. Postma, J.P.M., Berendsen, H.J.C. and Haak, J.R., *Faraday Symp. Chem. Soc.* 17, 55 (1982).
15. Wong, C.F. and McCammon, J.A., *J. Am. Chem. Soc.* 108, 3830 (1986).
16. Rao, S.N., Singh, U.C., Bash, P.A. and Kollman, P.A., *Nature (London)* 328, 551 (1987).
17. (a) Zwanzig, R.W., *J. Chem. Phys.* 22, 1420 (1954).
(b) Valleau, J.P. and Torrie, G.M., in *Modern Theoretical Chemistry* (ed. B.J. Berne), vol. 5, p. 169, Plenum, New York (1977).
(c) Valleau, J.P. and Whittington, S.G., in *Modern Theoretical Chemistry* (ed. B.J. Berne), vol. 5, p. 137, Plenum, New York (1977).
18. (a) Warshel, A., *Chem. Phys. Lett.* 55, 454 (1978).
(b) Warshel, A., *J. Phys. Chem.* 83, 1640 (1979).
19. (a) Warshel, A. and King, G., *Chem. Phys. Lett.* 121, 124 (1985).
(b) King, G. and Warshel, A., *J. Phys. Chem.* 91, 3647 (1989).
20. Warshel, A., Sussman, F. and King, G., *Biochemistry* 25, 8368 (1986).
21. Kuwajima, S. and Warshel, A., *J. Chem. Phys.* 89, 3751 (1988).
22. Carter, P. and Wells, J.A., *Nature (London)* 332, 564 (1988).
23. Tucker, P.W., Hazen, E.E., Jr. and Cotton, F.A., *Mol. Cell. Biochem.* 22, 67 (1978).
24. Cotton, F.A., Hazen, E.E., Jr. and Legg, M.J., *Proc. Natl. Acad. Sci. U.S.A.* 76, 2551 (1979).

25. Sepersu, E.H., Shortle, D. and Mildvan, A.S., *Biochemistry* 26, 1289 (1987).
26. (a) Åqvist, J. and Warshel, A., *Biochemistry* 28, 4680 (1989).
(b) Åqvist, J. and Warshel, A., *J. Am. Chem. Soc.* (in press).
27. (a) Price, P.A., *J. Biol. Chem.* 250, 1981 (1975).
(b) Moore, S. in *the Enzymes*. Vol. 14, p. 281 (1981).
28. Urry, D.W., *Proc. Natl. Acad. Sci. U.S.A.* 68, 672 (1971).
29. Bamberg, E. and Läuger, P., *Biochim. Biophys. Acta* 367, 127 (1974).
30. Eisenman, G. and Sandblom, J.P., in *Physical Chemistry of Transmembrane Ion Motions* (ed. G. Spach), p. 329, Elsevier, Amsterdam (1983).
31. Parsegian, A., *Nature (London)* 221, 844 (1969).
32. Levitt, D.G., *Biophys. J.* 22, 209 (1978).
33. Kim, K.S. and Clementi, E., *J. Am. Chem. Soc.* 107, 5504 (1985).
34. Etchebest, C., Ranganathan, S., and Pullman, A. *FEBS Lett.* 173, 301 (1984).
35. Jordan, P.C., *J. Phys. Chem.* 91, 6582 (1987).
36. (a) Åqvist, J. and Warshel, A., *Comm. Mol. Cell. Biophys.* 6, 91 (1989).
(b) Åqvist, J. and Warshel, A., *Biophys. J.* 56, 171 (1989).
37. (a) Deisenhofer, J., Epp, O., Miki, K., Huber, R. and Michel, H., *Nature (London)* 318, 618 (1985).
(b) Chang, C.H., Tiede, D., Tang, J., Smith, U., Norris, J. and Schiffer, M., *FEBS Lett.* 205, 82 (1986).
(c) Allen, J.P., Feher, G., Yeates, T.O., Komiya, H. and Rees, D.C., *Proc. Natl. Acad. Sci. U.S.A.* 84, 5730 (1987).
38. Creighton, S., Hwang, J.K., Warshel, A., Parson, W.W. and Norris, J., *Biochemistry* 27, 774 (1988).
39. Warshel, A., Chu, Z.T. and Parson, W.W., *Science* 246, 112 (1989).
40. Warshel, A., *Nature (London)* 260, 679 (1976).

DISCUSSION

DURUP - I feel there's a problem with your model for SN2 reactions in solution. You essentially are building diabatic states (your "resonant states") and then diagonalizing H to get an adiabatic curve. Now if you would do molecular dynamics and run a reactive trajectory, you would find a time of the order of 10^{-13} sec for this process, whereas the time necessary for the water molecules to rearrange is of the order of 10^{-11} sec. Thus I think the actual activation energy should be somewhere between the adiabatic one and the one you obtain from the crossing between diabatic states. I'm aware that you are not computing activation energies but free energies of reaction, but is it enough to remove the difficulty ?

WARSHEL - Sorry but my treatment which is quite complicated is at present the most rigorous treatment of the solute-solvent dynamic coupling. Please read J. Chem. Phys., 1988 and J. Am. Chem. Soc. 1988, 119, 5297. Not only that we include the solvent in the solute Hamiltonians but we developed an approach based on linear response theory to simulate the solvent and solute dynamics. We have a treatment for both the diabatic and adiabatic limit and the S_N2 is entirely within the adiabatic limit, also the solvent dynamics is in the 10^{-13} limit. Please read our detailed study of this in J. Am. Chem. Soc. 109, 715 (1987).

RIVAIL - The V.B. approach which you presented reminds me of a proposal of W. Miller to approximate the height of the potential energy barriers.

WARSHEL - I am not aware of the exact work but clearly more and more people are realizing the power of the simple diabatic representation.

PERAHIA - The Langevin dipole model misses the correlation effects between neighboring water molecules. A dielectric constant of 80 is obtained when such correlations are considered. What are the implications of this limitation of the model on your results ?

WARSHEL - In one of our recent version of the model we introduced the short range Kirkwood g factor in the Langevin model. This goes beyond the mean spherical

approximation. However, even with the original model that has dielectric constant of around 20 we are getting correctly the self energy of charges in protein.

WIPFF - How did you parametrize the parameter of Ca^{2+} used for modelling Staphylococcal Nuclease ?

WARSHEL - The parameter for Ca^{2+} are given in Aqvist and Warshel, Biochemistry, 28, 4680 (1989).

CHOJNACKI - Could you comment on the non-orthogonality problem of your empirical valence bond method ?

WARSHEL - Good question. In principle we consider our parameters to be obtained from the V.B. Hamiltonian after Löwdin orthogonalization procedure. Please see eq. (8) in J. Am. Chem. Soc. 110, 5297 (1988).

MODELLING THE ACETYLCHOLINE RECEPTOR (AChR) CHANNEL. ENERGY
PROFILES AND POINT MUTATIONS.

A. PULLMAN, S. FUROIS-CORBIN and A.M. ANDRADE

Institut de Biologie Physico-Chimique, 13 rue Pierre et Marie
Curie, Paris 75005 (France)

SUMMARY

A summary is given of the development of a molecular model of the inner wall of the AChR channel compatible with the presently existing structural knowledge (sequence, stoichiometry, etc) supplemented by conditions derived from recent labelling experiments and by theoretical calculations. Energy profiles are shown to explain the recently described effect of point mutations on the channel conductance.

INTRODUCTION : THE MODEL

In the currently prevalent concept of the AChR channel (1), five pentagonally disposed subunits ($\alpha \beta \alpha \gamma \delta$) are believed to participate in the transmembrane structure by four hydrophobic helices MI-MIV and to contribute at least one helix, MII, to the inner wall of the channel. Labelling experiments (2-5) with competitive blockers (NCB), identifying the labelled residues as homologous serines in the five MII helices, have suggested (6-7) that these residues face the interior of the pore, thereby fixing the orientation of the MII's with respect to the central axis.

Adopting these premisses and using theoretical calculations, a molecular model of the channel was built as follows : it was first shown in a theoretical study (8) that, in order to satisfy the suggestion (3) that the high affinity site for the NCB chlorpromazine lies at or near the level of the labelled serines on the axis of quasi-symmetry of the receptor and at minimum distances from all five chains, consecutive MII helices must be laterally in contact at this level rather than separated by another helix. It was shown furthermore (9) that this hypothesis allows the calculation of the minimal distance of the helices in this region. Moreover, adopting the reasonable assumption that the large permeant ions or NCB's must diffuse through the upper part of the channel, it was noted that the presence of bulky internal

side-chains in this upper part of the inner wall would prevent the diffusion if the five MII helices were parallel : using the sequences of the MII's (figure 1) and assuming the Ca 's of the labelled serines (numbered 8) to face the center of the pore, the "helix wheels" request (8) that the other Ca carbons situated on the interior wall are those of residues 1, 4, 12, 15 and 19.

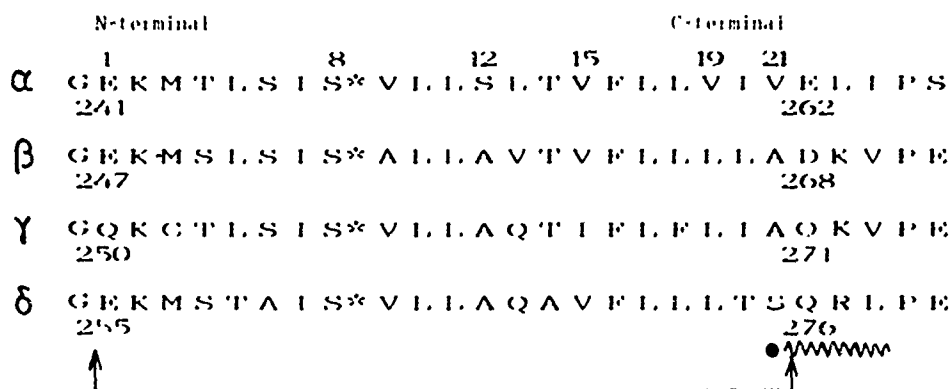


Fig. 1. The aligned sequences (standard numbering as indicated) of the MII segments of *Torpedo Marmorata* and *Torpedo Californica* (18) within the limits (between arrows) discussed in ref. 8 and 9 ; (*) : homologous serines labelled by $^3\text{[H]CPZ}$ and $^3\text{[H]TPMP}$. For practical reasons a simplified numbering is indicated on the top line.

Residues 15 and 19 which are in the upper internal part of the pore are bulky valines, isoleucines and leucines. Conformational optimizations and model building in this part of the helices allowed the determination of the minimal approach of the MIIs at this level compatible with the passage of chlorpromazine (9). This condition, together with that imposed at the level of the labelled serines impose to the channel wall a truncated conical shape (see figure 2). The resulting tilt of the helices was calculated to be 7 degrees with respect to the central vertical axis of the pore. This model leaves small gaps between adjacent MII's in the upper part of the channel wall, which can, however, be easily closed by contact with another helix of each subunit, either MI or MIII (see ref. 9 for a discussion).

An important feature of the model, discussed in ref. (8), is the inclusion in the α -helical portion of the MII segments of the

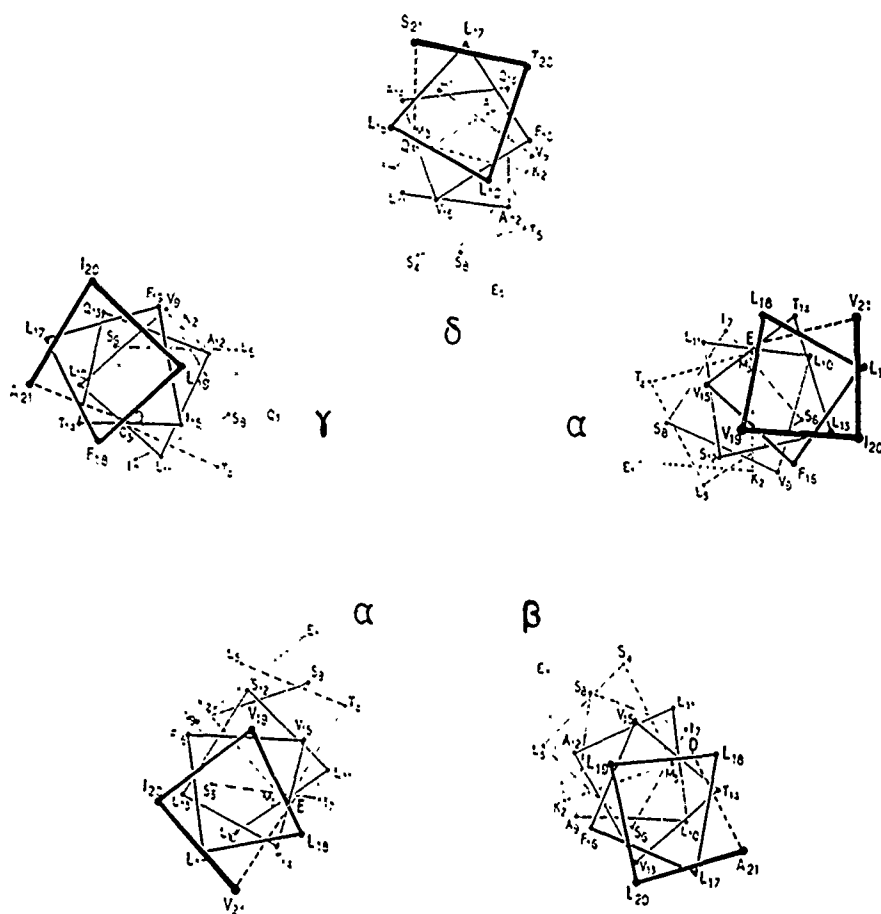


Fig. 2. A view from the synaptic side of the disposition of the five MII helices around the pore. The points labelled by letters represent the location of the $C\alpha$ carbons of the successive residues in the α -helical structure. Numbering and amino acid code as in figure 1. As discussed in (9) the last residue is the 21st one and residue 22 (not numbered) is outside the helical stretch.

charged glutamates (Gln in γ) situated two residues below the sulfur amino-acids generally considered as the N-terminal groups. Energy calculations of the energy profile of the largest permeant ion dimethyldiethanolammonium (DMDEA) showed that this inclusion is fundamental in determining the possibility of exit of the ions (8). This feature results from the α -helical structure which, together with the internal positions of the labelled serines imposes that the negatively charged glutamates (Gln in γ) face also the interior of the pore while the adjacent positively charged lysines point to the external side of the helices, thus do not annihilate the favorable action of the negative residues. These conclusions, hence the underlying hypothesis, have been strikingly confirmed by the site-directed mutation experiments of ref. (10) which showed, in particular, that mutating the negatively charged residues considered above has a much stronger effect on the channel conductance than mutating the other anionic residues found between MI and MII or between MII and MIII, while mutating the positively charged residues following the Glu(Gln) positions leads to no significant change in channel conductance. The dominance of the effect of the negative residues at positions 1 over that of the residues at positions 22 (see fig. 1) is in very good agreement with the fact underlined in (9) that the upper limit of the MII must be placed at residue 21, owing to the presence of a proline four residues ahead in the sequence.

A recent calculation (11) of the energy profile for a sodium cation in the model has confirmed and extended these previous findings. It has furthermore indicated the role, in the profile, of the other helices of the subunits. Thus, it was shown that when only the five MII helices are used to calculate a profile, the energy (even though negative, thus favorable) is becoming less and less favorable from top to bottom of the pore, a situation disadvantageous for the passage of the cation. This reflects the fact (7,9,11) that the variations of the binding energy are determined by the variations of its electrostatic component. Thus due to the large dipole moment (about 74 Debye units) of each MII α -helix, the conical bundle made of five nearly parallel MII's presents a very large dipole moment with the negative pole oriented towards the synaptic side and the positive pole towards the cytoplasmic side, a situation which would be unfavorable for the crossing of a positive cation if the MII's were alone. But in fact, in the actual situation, each MII belongs to a subunit which

comprises four helices, alternately down and up from the synaptic side to the cytoplasm, thus with alternately oriented dipole moments. Thus, even though the MII's are the closest to the pore, the unfavorable effect of their parallel dipoles on the passage of a positive cation is considerably reduced by the presence of the other helices of the receptor : it was shown (11) that the effect in this respect of the three supplementary helices in each subunit could be mimicked by supplementing the cone of MII's by another cone of analogous α -helices antiparallel to the MII's and placed in contact with them just behind as indicated in figure 3 (see ref. 11 for the technical details).

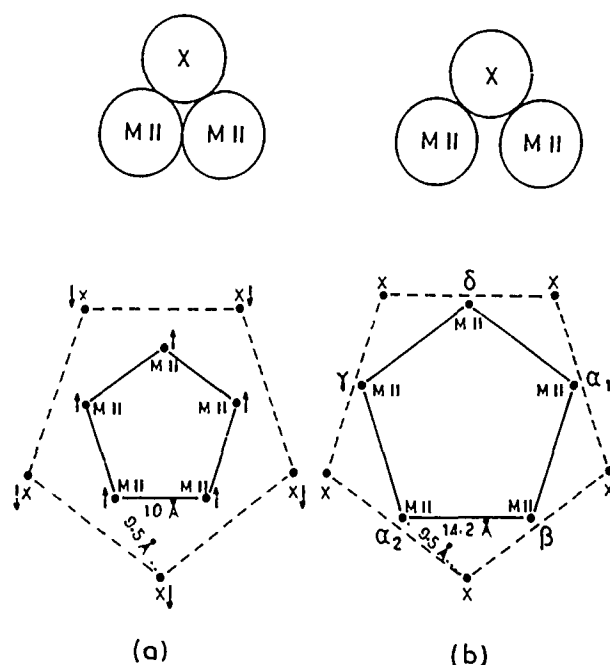


Fig. 3. Schematic view from the synaptic side of the mutual disposition of two adjacent MII helices and of the five helices X surrounding them in (a) lower part and (b) upper part of the conical model of the channel. Arrows indicate the orientation of the helices, from their N- to their C-terminal.

ENERGY PROFILES IN THE MODEL

The "energy profile" of Na^+ computed in the double cone so defined indicated that the interaction energy E of Na^+ with the whole "double" bundle (ten helices) is negative, thus favorable, and presents a favorable evolution from top to bottom of the channel (figure 4) a result of the now favorable evolution of its electrostatic component.

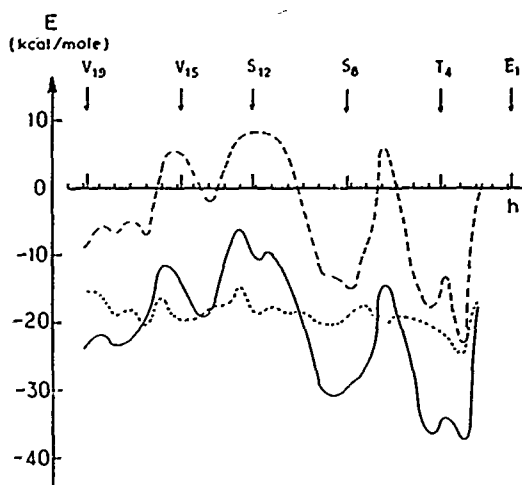


Fig. 4. Energy profile of Na^+ and its components in the double cone of fig. 2 : (---) total interaction energy E ; (- - -) electrostatic component of E ; (.....) polarization component of E . The Lennard-Jones component is very small, thus omitted in the figure. The scale h indicates in Å the height of Na^+ in the channel (from 30, left : synaptic side to 0, right : cytoplasmic side). The arrows indicate the height of the $\text{C}\alpha$ of the residues facing the interior of the pore in MII α (see fig. 1 for the homologs).

It is interesting to consider the effect on such a profile of representative mutations experimentally tested (10) in the anionic ring situated at the bottom of the MII's. This is illustrated below on two examples, namely the substitution of Glu 255 in δ by Gln and of Glu 241 in α by Asp (residues 1 in the numbering of figure 1).

The calculation of the energy profile was done as in (8,9,11-13) by optimization of the energy of interaction (13) of the ion with the whole channel in successive planes perpendicular to the central axis, regularly and closely spaced. The "energy profile" is the plot of this interaction energy as a function of the progression of the ion. As in (8) the channel is maintained rigid. The computations were performed in the double cone defined in ref. 11 as recalled above.

The introduction of the mutation is made as follows : according to energy optimizations carried out in the model channel (8,9) Glu 1 and Lys 2 form a salt bridge in α , β , δ ; Gln 1 and Lys 2 in γ are H-bonded. In keeping with this standpoint, for the substitution of Glu by Gln in MII δ , a conformation was adopted where the new Gln forms with Lys 2 a hydrogen bond analogous to that of Gln 1 and Lys 2 in MII γ . For the substitution of Glu by

Asp in MII α , a real salt bridge like that formed by Glu and Lys in the unsubstituted channel cannot be achieved, owing to the shorter chain of the aspartate, but it appears reasonable to adopt a structure where one of the negatively charged carboxyl oxygens of Asp is directed towards the positively charged end group of Lys in an elongated bridge.

For simplicity we call W (wild-type) the double cone of the model where the sequences of the five MII's are those of Fig. 1, E δ Q the same double cone where Glu 1 in MII δ is replaced by Gln and E α D the double cone where Glu 1 in both MII α 's is replaced by Asp.

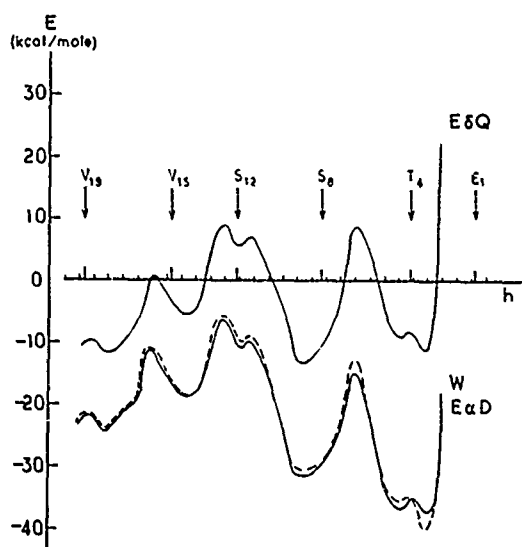


Fig. 5. Energy profiles of Na^+ in the double cone of figure 2. The curves W and E δ Q in full line are the profiles in the wild-type channel and in the E δ Q mutant respectively ; the dotted line is the profile in the E α D double mutant. Scale as in fig. 4.

Fig. 5 compares the profiles calculated in W, E δ Q and E α D respectively. The disposition of the curves W and E δ Q shows the effect of the involved substitution on the crossing of the ion : the energy of interaction of Na^+ with the bundle is considerably affected. The modification occurs not only in the lower part (close to the modified residue) but all the way through the channel : the energy in E δ Q is less negative, thus less favorable, than in the wild type. As in W (11), the variations of the total interaction energy of Na^+ with E δ Q are governed by the variations of the electrostatic component of this energy, by far less favorable in E δ Q than in W, the other components of the

interaction energy being not affected (Fig. 6 to be compared with Fig. 4).

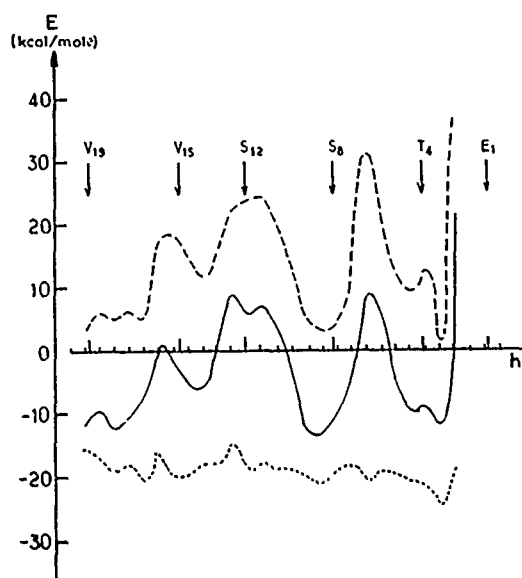


Fig. 6. Energy components of the profile in the mutant E8Q :

----- total

- - - electrostatic component

..... polarization

As in W the Lennard-Jones component, small everywhere, is not given. Scale as in fig. 4.

If we consider, on the other hand, the profile obtained after mutating Glu to Asp in the two α MII's, we observe that it is nearly undistinguishable from the profile obtained in the wild type.

Thus on the basis of the energy results, the transit of the ion should be appreciably less favorable in the E8Q mutant than in the wild type whereas it should be about as favorable in the E8D mutant as in the wild type : these results appear in remarkable agreement with the observations of ref. (10) where indeed the E8Q mutation lowered the channel conductance by about one half, whereas the E8D mutation had very little effect.

The reasons of the variations observed in the global energy profiles upon mutating towards E8Q can be still better understood upon considering the trajectory (Fig. 7) followed by the ion together with the values of its interaction energies with each helix individually along its path (Fig. 8). Very strikingly, the successive optimal points reached by the ion in W and in E8Q are essentially the same (whatever the starting point chosen for the optimization). Moreover, descending towards the N-terminal, these

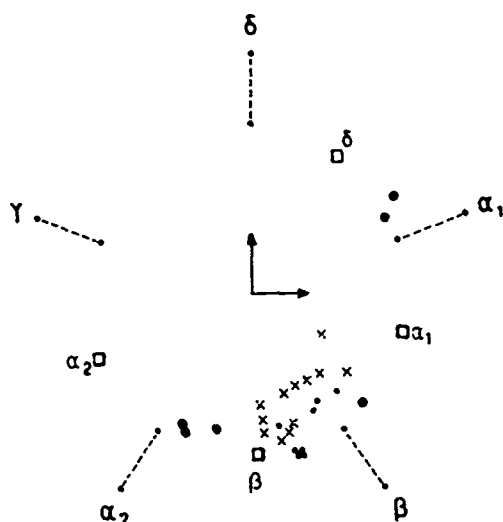


Fig. 7. The optimal points reached by Na^+ in the successive planes studied.

- Upper part of the channel down to the level of the $\text{C}\alpha$ of residues 16 (according to the simplified numbering in Fig. 1).
- From the level of the $\text{C}\alpha$ of residues 15 down to the level of the $\text{C}\alpha$ of residues 11.
- x Next portion of the channel down to the N-terminal of MII's.
- Joins the traces of the axes of the MII's at the bottom and the top of the cone.
- position of the free oxygen of Glu.

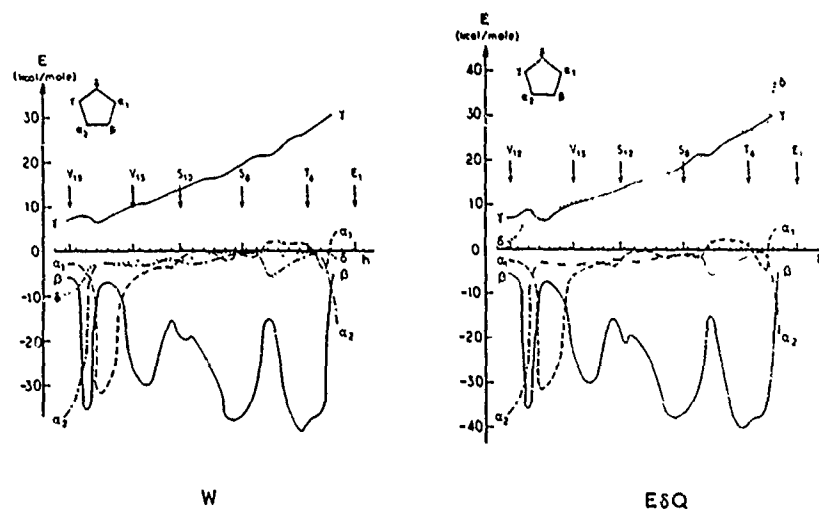


Fig. 8. Contribution of each MII helix to the energy profiles in W and in EsQ as indicated: the curves α_1 , α_2 , β , γ , δ correspond to the interaction energy of Na^+ with the respective helices (order indicated in insert).

points rapidly cluster in the neighborhood of MII β : in the wild-type channel, the cation is more and more attracted, all along its progression from top to bottom of the channel, towards the region located near MII β where the conjunction of the attractive effects of the four glutamates present is maximum. The presence of a Gln instead of a Glu in MII γ makes the region surrounding this helix less favorable. The individual interaction energies of Na⁺ with the different helices (Fig. 8,W) reflect its proximity to each of them respectively : note the most favorable interaction with β , the favorable but smaller (except at the extremes) interaction with α_1 and α_2 , the favorable but constantly small interaction with δ and finally, the unfavorable interaction with γ , increasing towards the bottom of this helix where no negative charge is present to attract the ion.

When Glu in δ is mutated to Gln, the convergence of the attraction remains in the neighbourhood of β and the optimal positions remain essentially the same with interactions with β , α_1 , α_2 similar as before. But now, the interactions with δ and with γ are both unfavorable (Fig. 8,E δ Q), more so towards the bottom of the helices which both lack a negative charge. Hence an overall less favorable global energy.

When Glu is mutated into Asp in the two α 's the situation is very similar to that of the wild type, both in the path of the ion and in the individual interactions : in the mutated channel, in spite of the shorter length of the Asp side chains, the field which they create on the ion is still strong and the convergence of the effects of α Asp, β Glu, α Asp is clearly similar to that of α Glu, β Glu, α Glu.

CONCLUDING REMARKS

The energy profiles calculated for the mutants E255Q in δ and E241D in α of the AChR channel using the proposed model of the channel inner wall account remarkably well for the variations in the observed conductance with respect to that of the wild-type. The analysis of the data confirms and specifies the role proposed (8) for the negatively charged residues at the N-terminal of the MII helices. The agreement of the profiles calculated within the model with the conclusions of the conductance experiments appears as a confirmation of the essential structural hypotheses of this model. Although the important role of the N-terminal negative residues seems well established both by theory (8,11) and

experiment (10), it is possible that they depart somewhat from the bridge structures found by energy optimization in vacuo (14). The conjunction of theoretical calculations and experimental observations on other mutants (10,15) should help a further refinement of the model.

Until a crystal structure at high resolution of the transmembrane part of the receptor becomes available, progress towards the understanding of its functioning relies heavily on further modelisation in close contact with experiment. We hope that the present model can serve as a starting point.

REFERENCES

- 1 See for instance Transport through Membranes : Carriers, Channels and Pumps, in: A. Pullman, J. Jortner and B. Pullman (Eds.), Kluwer Academic Publishers, Dordrecht, Holland, 1988.
- 2 J. Giraudat, M. Dennis, T. Heidmann, J.Y. Chang and J.P. Changeux, Proc. Natl. Acad. Sci. USA, 83 (1986) 2719-2723.
- 3 J.P. Changeux and F. Revah, TINS, 10 (1987) 245-249.
- 4 F. Hucho, Eur. J. Biochem., 158 (1986) 211-226.
- 5 F. Hucho, W. Oberthur and F. Lottspeich, FEBS Lett., 205 (1986) 137-142.
- 6 J. Giraudat, M. Dennis, T. Heidmann, P.Y. Haumont, F. Lederer and J.P. Changeux, Biochemistry 26 (1987) 2410-2418.
- 7 F. Revah and J.P. Changeux, in: A. Pullman, J. Jortner and B. Pullman (Eds.), Transport through Membranes : Carriers, Channels and Pumps, Kluwer Academic Publishers, Dordrecht, Holland, 1988, pp. 321-335.
- 8 S. Furois-Corbin and A. Pullman, in: A. Pullman, J. Jortner and B. Pullman (Eds.), Transport through Membranes : Carriers, Channels and Pumps, Kluwer Academic Publishers, Dordrecht, Holland, 1988, pp. 337-358.
- 9 S. Furois-Corbin and A. Pullman, Biochim. Biophys. Acta, in press.
- 10 K. Imoto, C. Busch, B. Sakmann, M. Mishina, T. Konno, J. Nakai, H. Bujo, Y. Mori, K. Fukuda and S. Numa, Nature, 335 (1988) 645-648.
- 11 S. Furois-Corbin and A. Pullman, FEBS Lett., 252 (1989) 63-68.
- 12 C. Etchebest and A. Pullman, J. Biomol. Struct. Dyn, 4 (1986) 805-825.
- 13 S. Furois-Corbin and A. Pullman, Biochim. Biophys. Acta, 860 (1986) 165-177.
- 14 S. Furois-Corbin, A.M. Andrade and A. Pullman, in preparation.
- 15 R.J. Leonard, C.G. Labarca, P. Charnet, N. Davidson and H.A. Lester, Science, 242 (1988) 1578-1581.

DISCUSSION

WIPFF - The interaction energy of Na^+ with the bundle built up from poly Ala neutral residues is quite large and attractive (about - 50 kcal/mole) but is only about half of the dehydration energy of the ion. Where would you suggest these additional required 50 kcal/mol could come from for passing through the Channel in water ?

A. PULLMAN - The calculation of the profile for a bare Na^+ in the package of five poly Ala was done in order to explore whether a bundle of entirely hydrophobic non polar helices could accommodate a cation in spite of the absence of polar residues inside the pore, the common belief at the time of the calculation (BBA, 1986, ref. 13 of manuscript) being that polar and even ionized residues were necessary. We did not pretend to touch the problem of entrance or/and desolvation.

I mentioned the role of polar (non ionized) residues when present (like serines which are often there in small number). They provide extra-binding energy or/and can help local disturbance of water structure (J. Mol. Str. and Dynamics 1987, 4, 589-598). Entrance desolvation is probably progressive and helped by some polar (often ionic) groups.

PERAHIA - Your model does not contain water molecules inside the channel which may have important screening effects. The electrostatic energy variations will be much lower if you take into account screening effects, and the electrostatic energy profile may tend to be different.

A. PULLMAN - We have not yet introduced water explicitly. However (particularly for the last problem considered) we have done the same calculations with different (tentative) value of the dielectric constant : although, truly, it affects the values of the energies, hence the profile, by screening the electrostatic component, it did not affect the overall results concerning the location of the trajectory of Na^+ and the reasons for it. It must be added that various experimental evidences indicate that only a few water molecules are in the narrow (low) part of the AChR channel, hence a relatively small screening only can occur for the effect of the four glutamates which we advocate.

MAROUN - In the energy profile of the passage of the Na^+ ion from the top to the bottom of the channel, there is an energy barrier of about 15 Kcal around residue 58. How can the ion, then, go through this middle-of-the channel barrier ?

A. PULLMAN - As I said the model does not include complete optimization (of the side chain conformations or of the bundle itself) neither without nor with the ion. Such optimizations can very easily lower or abolish small barriers. This has been explicitly shown in our early calculations on the model bundles with Ala, Leu, Ser and Glu (See our BBA papers in 1986-1987-1988).

BUCKINGHAM - It is thought that entropic effects play an important role in hydrophobic interactions. Do you think they are important in transport through membranes ?

A. PULLMAN - This is a question we did not touch for the moment.

ANALYSING AND MODELLING THE DEFORMATION OF PROTEIN SECONDARY STRUCTURES

C. ETCHEBEST

Laboratoire de Biochimie Théorique, Institut de Biologie Physico-Chimique, 13, rue Pierre et Marie Curie, Paris 75005 (France)

ABSTRACT

A new method able to perform controlled deformations of oligopeptides with the help of geometrical constraints is described. The deformation energy for two types of deformation applied to α -helices with three different amino acid sequences has been computed. The results show that this approach, coupled with the P-Curves method for analyzing polypeptide structure, is particularly appropriate for producing global modifications of conformation.

INTRODUCTION

Secondary structures, most notably, α -helices and β -sheets form a considerable part of most protein conformations. However, these structural elements rarely approach perfect helicoidal symmetry and their participation in the overall conformation often involves considerable deformations. In order to understand the energetic nature of these effects we have developed a technique that is able to produce easily controlled deformations of basic structural elements and enables us to perform theoretical calculations of deformation energy. In order to study the influence of the peptide sequence and to clarify, in particular, the role of proline, we have investigated three different α helicoidal sequences of 21 amino acids, (Ala)₂₁, (Val)₂₁ and (Ala)₁₀-Pro-(Ala)₁₀.

METHOD

In order to apply constraints between any 2 chosen peptide units i and j we firstly set up an axis system U, V, W centered at point P which indicates the location of the helical axis system in the starting conformation (Fig. 1). The four variables (2 translations and 2 rotations) necessary to fix this axis system

have already been described in our previous work concerning the P-Curves algorithm (ref. 1-2).

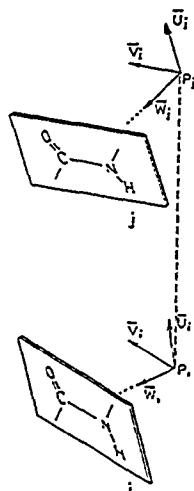


Fig. 1 : Definition of the axis systems for peptide i and j.

These axis systems can now be used to construct various geometrical constraints :

- 1) Stretch or lateral dislocation between peptide i and peptide j : $(\bar{P}_i - \bar{P}_j) \cdot \bar{X}_i$, where \bar{X}_i is \bar{U}_i or \bar{V}_i or \bar{W}_i .
- 2) Helical axis bend between peptide i and peptide j : $\bar{U}_i \cdot \bar{U}_j$.
- 3) Bending Direction : $\bar{U}_j \cdot \bar{V}_i / |(\bar{U}_j - (\bar{U}_i \cdot \bar{U}_j) \bar{U}_i)|$, the sign being defined by : $(\bar{U}_j \times \bar{V}_i) \cdot \bar{U}_i$
- 4) Helical twist Angle : $\bar{V}_i \cdot \bar{V}_j$ (maintaining \bar{U}_i and \bar{U}_j colinear) the sign being defined by : $(\bar{V}_i \times \bar{V}_j) \cdot \bar{U}_i$

These geometrical constraints are imposed during the minimisation process of the conformational energy whose the formula is given by five terms : electrostatic, short range repulsion, dispersion, dihedral angle rotation and valence angle distortion (see ref. 3 for details). For this purpose analytic first derivatives of the the active constraints are calculated and used in conjunction with the energy and its derivatives by an advanced variable metric constraint minimisation routine. In the case of the polypeptide containing a proline, the computations are made allowing the proline ring to be flexible, assuring the ring

closure again with the help of a geometrical equality constraint (ref. 4).

The starting conformations for (Ala)₂₁ and (Val)₂₁ helices were obtained with a standard set of ϕ , ψ and ω angles (ref. 5) respectively -57° , -47° , 180° , and with χ angles equal to 180° . For the proline residue, the values were -74.2° , -47.0° , 180.0° and the χ angles for the ring were taken as the most statistically probable conformation (ref. 6). The deformation energy is given by the difference between the optimal conformational energies of the constrained structure and the freely minimised reference structure.

During optimisation, the axis systems defined above are locked to their respective peptides. Once the final structure is obtained a P-Curves analysis is made to find the optimal helical description of the now distorted oligopeptide. This analysis can lead to changes in the desired constraint geometry (e.g bending angle) but these changes are, in general, negligible.

Two types of deformation have been studied for the three sequences mentioned above : stretch and bending angle (for two directions). The constraints have been imposed between residues 2 and 20 of the sequence.

RESULTS

Before studying the effect of the various deformations, we have analysed with the P-Curves algorithm the optimised conformations of the three sequences. The results show that the (Ala)₂₁ and (Val)₂₁ are linear, whereas the sequence containing a proline is naturally curved, the corresponding bending angle between residues 2 and 20 being 17° .

Let us now consider the results obtained when the structures are stretched. In fig 2 we present the energy variations for the sequences (Ala)₂₁, (Val)₂₁ (curve A and B respectively) with respect to the stretching. In both cases, the deformation energy increases rapidly until roughly 40 Kcal/mol for 4 Å elongation. This energy loss is due to the progressive disruption of the ensemble of hydrogen bonds. Nevertheless it may be noted that the breaking is smooth, hydrogen bonds being lengthened and becoming more bent. In some cases this is compensated by the formation of new hydrogen bonds of the type 1-3 (3_{10} helix), situated at the end of the oligopeptides. This situation is more marked for

(Ala)₂₁ as indicated by the inflexion point occurring between 1.5 and 2.5 Å on curve A. This may be explained through a larger flexibility due to the alanine side chains which occupy less volume than valine.

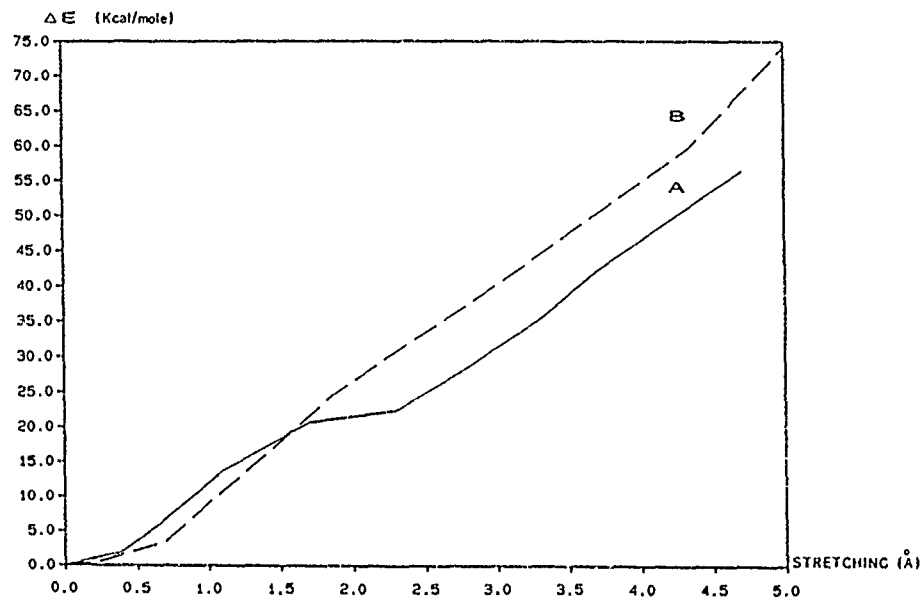


Fig. 2 : Deformation energy with respect to stretching. Curve A : (Ala)₂₁, curve B : (Val)₂₁.

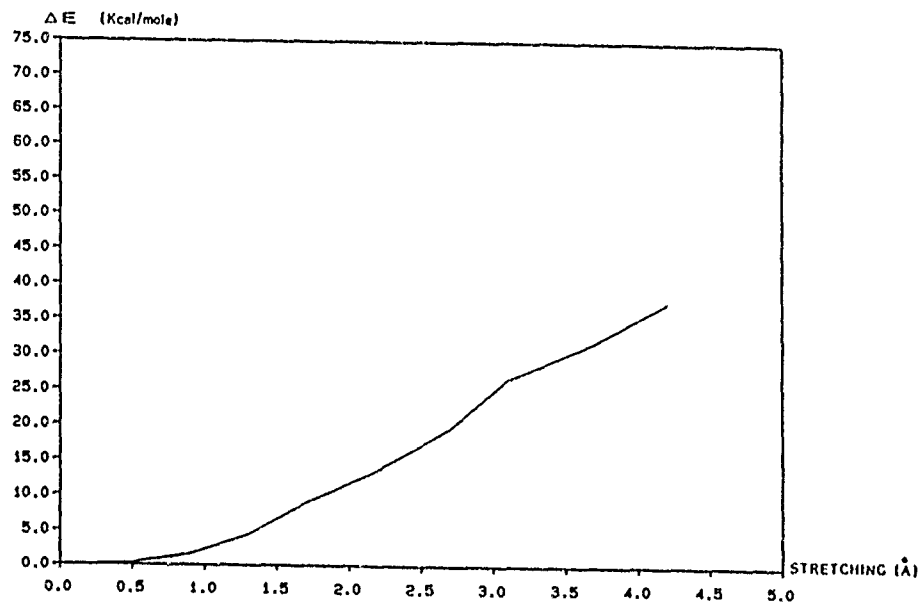


Fig. 3 : Deformation energy with respect to stretching. for (Ala)₁₀-Pro-(Ala)₁₀.

The results for the third sequence containing a proline residue are shown in fig. 3. The energy variation is smaller than that for (Ala)₂₁ or (Val)₂₁. It is clear in consequence that the presence of a proline residue facilitates this deformation. The mechanism involved in stretching is the same : hydrogen bonds of 1-3 type are formed, the location of these hydrogen bonds is however not at the end of the structure but rather in the immediate vicinity of the proline. Moreover we have observed a modification of the global curvature of the structure.

Concerning the bending angle constraint, we have studied two perpendicular directions for the three sequences. Fig. 4 illustrates the results for (Ala)₂₁, the two curves representing the two directions of curvature. The differences between the two curves indicate an anisotropy of bending amounting to almost 2 Kcal/mol. This anisotropy may be attributed to the angular distribution of the hydrogen bonds within the α -helical structure of the oligopeptide.

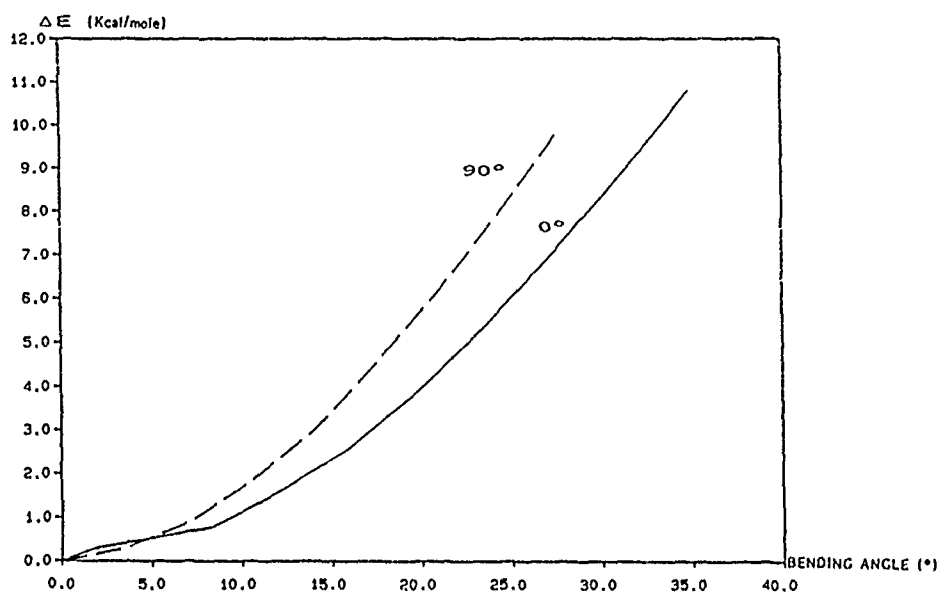


Fig. 4 : Deformation energy with respect to the bending angle in two perpendicular directions for (Ala)₂₁.

The form of both curves is nevertheless similar and quadratic. The mean force constant computed is 1.25×10^{-2} Kcal/mol/Å². The energy loss reaches roughly 10 Kcal/mol for a bending angle of 30-35°. Bending occurs through the progressive

deformation of the hydrogen bonds, but, in contrast to the stretching deformations, the number and the type of these bonds are conserved.

The results for (Val)₂₁ are shown in fig. 5 and are globally similar to those for (Ala)₂₁ (mean force constant for bending = 1.1×10^{-2} Kcal/mol/Å²). It can thus be remarked that the larger valine side chains appear to play a smaller role in affecting bending than was the case for stretching.

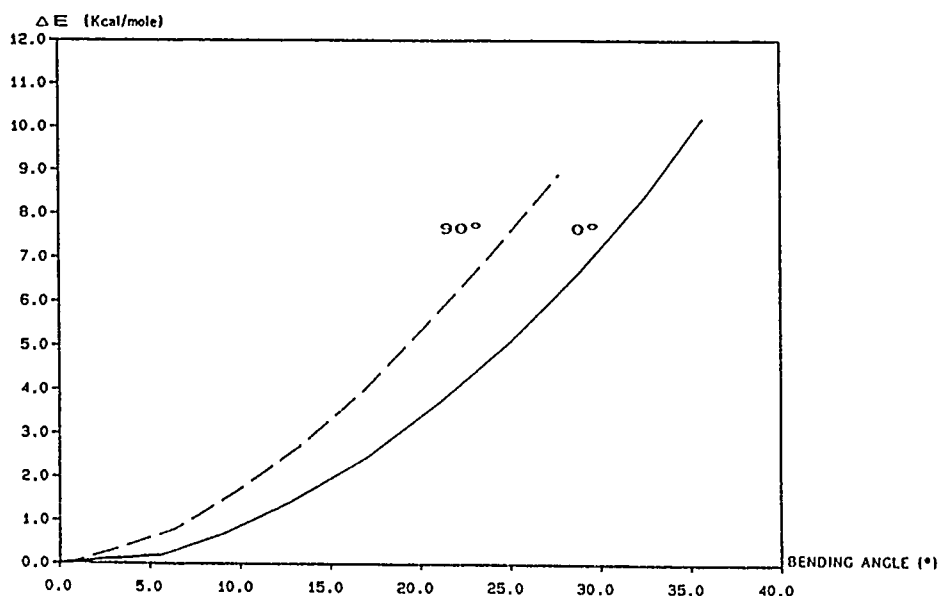


Fig. 5 : Deformation energy with respect to the bending angle in two perpendicular directions for (Val)₂₁.

The final results concerning the sequence (Ala)₁₀-Pro-(Ala)₁₀ are given in fig. 6. As a first remark it may be noted that the deformation energy is only 4-7 Kcal/mol for 35° of bending, the value depending on the direction. These values are much smaller than those obtained with the other sequences. This result indicates that the presence of the proline facilitates α -helix bending. Nevertheless, as the results show its bending is anisotropic. A second difference resides in the shape of the bending curves : the deformation energy in the two directions is not quadratic. Despite these differences, the mechanisms involved are the same described previously and consist of the progressive deformation of the hydrogen bonds.

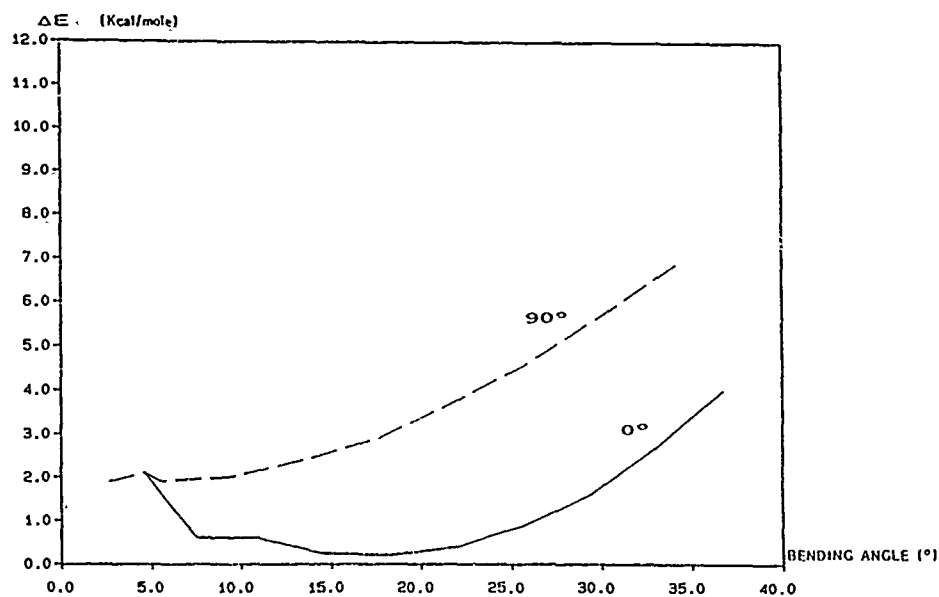


Fig. 6 : Deformation energy with respect to the bending angle in two perpendicular directions for $(\text{Ala})_{10}\text{-Pro-(Ala)}_{10}$.

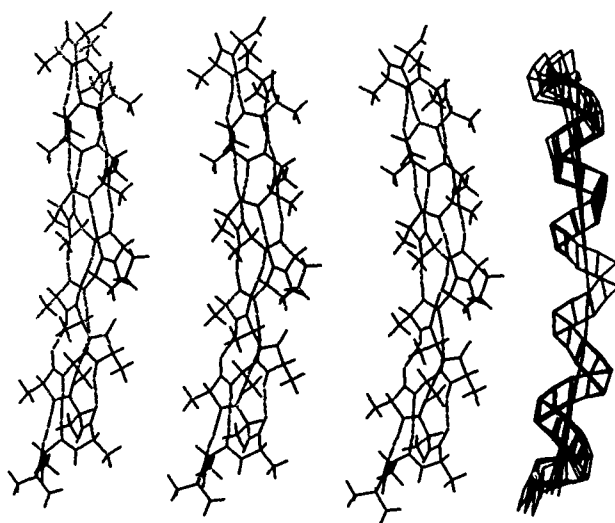


Fig. 7 : $(\text{Ala})_{10}\text{-Pro-(Ala)}_{10}$: optimized structures for 3 different bending angles and the corresponding helical axis.

The effect of the geometrical bending constraint can be more easily visualised with the help of Fig. 7. In this figure, we show the molecular diagrams of the optimised structures computed for three different bending angles for the sequence (Ala)₁₀-Pro-(Ala)₁₀. The hydrogen bonds are indicated by dotted lines. The corresponding helical axis and the peptide backbone (represented with a ribbon), computed with the P-Curves method, are indicated on the right part of the figure.

CONCLUSIONS

The results obtained in this study have brought to light several properties of the α -helix :

- 1) in all cases the energy increases rapidly with stretching.
- 2) for (Val)₂₁ and (Ala)₂₁, the variation of energy with respect to bending is quadratic.
- 3) When a proline is included in the helix, bending becomes much easier, but shows a marked anisotropy.

This simple study demonstrates the use of appropriately defined geometrical constraints within an internal coordinate formalism. Future applications of this technique will allow us to define sequence dependant flexibilities for polypeptide structural motifs which can subsequently serve for the construction of simplified, large scale models of protein architecture.

REFERENCES

- (1) H. Sklenar, C. Etchebest and R. Lavery, Describing protein structure : a general algorithm yielding complete helicoidal parameters and a unique overall axis, *Proteins : Structure, Function and Genetics*, in press.
- (2) K. Zakrzewska and R. Lavery, Describing protein conformation : a new mathematical approach, this volume.
- (3) R. Lavery, H. Sklenar, K. Zakrzewska and B. Pullman, The flexibility of the nucleic acids : (II) the calculation of internal energy and applications to mononucleotide repeat DNA, *J. Biomol. Struct. Dynam.*, 3 (1986) 989-1014.
- (4) R. Lavery, I. Parker and J. Kendrick, A general approach to the optimization of the conformation of ring molecules with an application to valinomycin, *J. Biomol. Struct. Dynam.*, 4 (1986) 443-462.
- (5) S. Arnott and S.D. Dover, Refinement of bond angles of an α -helix, *J. Mol. Biol.*, 30 (1967) 209-212.
- (6) J.W. Ponder and F.M. Richards, Tertiary templates for proteins, Use of packing criteria in the enumeration of allowed sequences for different structural classes, *J. Mol. Biol.*, 193 (1987) 775-791.

MONTE CARLO FREE ENERGY CALCULATIONS IN CONFORMATIONAL STATISTICS OF POLYPEPTIDE CHAINS.

S. PREMILAT

Laboratoire de Biophysique Moléculaire - UA CNRS N° 494 - Université de Nancy I
B.P. N° 239 - 54506 Vandoeuvre les Nancy (France).

SUMMARY

The partition of the total conformational energy of a chain molecule into short-range and long-range interactions allowed us to propose Monte Carlo procedures based on the simultaneous use of statistical weight matrices for short-range terms and importance sampling for the long-range interactions. Such an approach gives also a way of defining reference states of the molecular system. Therefore, it becomes possible to study the behavior of chain molecules under more complex conditions. The free energy and entropy variations due to long-range interactions are then calculated. Application of the present calculation methods are made on molecular models of peptidic hormones.

INTRODUCTION

Excluded volume, solvent effects as well as electrostatic interactions are mainly responsible for the folding of polypeptide chains. However these long-range interactions introduce major difficulties (refs. 1,2) because they prevent us from expressing the conformational energy as a sum of terms associated to each chain unit or to first neighbor pairs. Moreover, beside the commonly evaluated energy of configuration, the stability of macromolecules is also dependent on entropic contributions or, more generally, on the free energy of the system. But, there is generally no direct access to such quantities especially if the molecular system is subject to solvent effects.

The partition of the total configurational energy of a chain molecule into a part associated to short-range interactions and another one to the long-range interaction energy permits (ref. 3) to define reference systems for the chain molecules since when only short-range interactions are considered, all the statistical properties of the system are calculable. These reference systems are then used (ref. 4) for the calculation of entropy and free energy variations due to the effects of long-range interactions on the molecular chain conformation. Models of polypeptide chains and particularly of peptidic hormones are taken as examples of application of the proposed methods of calculation.

METHODS

Conformational energy

Atomic coordinates of polypeptide chain conformations are computed using

standard bond lengths and bond angles (ref. 5). The different residues of the peptide chain are classified, as described previously (ref. 3), into five types: GLY, PRO, ALA (for amino-acids with a side chain), a N-terminal unit and a C-terminal one. Every residue in a polypeptide chain has a definite type which corresponds to a precise energy map. The different maps are divided into regions or unit states.

The conformational energy of a polypeptide chain of n units is defined as the sum of all pairwise nonbonded (6-12 potentials) and electrostatic interactions between atoms plus torsional energy. Parameters and expressions of energy contributions presently used are already described elsewhere (ref. 5).

Let us thus consider the following expression of the total energy of a polypeptide chain in configuration η :

$$E_T^\eta = \sum_{i=1}^n E_i^t + \sum_{i=2}^n E_{i-1,i}^{r,t} + E_L^\eta = E_S^\eta + E_L^\eta \quad (1)$$

E_i^t corresponds to the contributions to the configurational energy from interactions between atom pairs within a unit (i) at the point t of its (φ, ψ) map. Terms $E_{i-1,i}^{r,t}$ are related to interactions between atoms located in two nearest-neighbor units in conformations defined by the points r and t of their respective maps. The two summations constitute the short-range interactions energy E_S whereas E_L represents the energy of long-range interactions between atom pairs situated in units more distant along the chain than first neighbors.

Chain with first neighbor interacting units

For molecular chains subject only to short-range interactions, the configurational partition function can be written as (with the Boltzmann constant k_B and temperature T) :

$$Z_S = \sum_{(C)} \sum_{\eta \in C} \exp(-\beta E_S^\eta) \quad ; \quad \beta = 1/k_B T \quad (2)$$

(C) is the set of chain conformations defined by all the different sequences of map regions which can be given to the chain units whereas the subset C is defined by giving a precise map region to every chain unit. The configuration η corresponds to a sequence of (φ, ψ) points taken in these precise regions. If there are n_ℓ points (φ, ψ) in the region ℓ given to unit (i) and n_k in region k allocated to ($i-1$), we have the following averages :

$$\alpha_{\ell,i} = \frac{1}{n_\ell} \sum_{t \in \ell} \exp(-\beta E_i^t) \quad \text{and} \quad \omega_{i-1,i}^{k,\ell} = \frac{1}{n_\ell n_k} \sum_{r \in k} \sum_{t \in \ell} \exp(-\beta E_{i-1,i}^{r,t}) \quad (3)$$

By substitution of these averages taken over map regions for the corresponding terms in $\exp(-\beta E_S^n)$, we can replace this last expression by :

$$\exp(-\beta E_S^C) = \prod_{i=1}^n \alpha_{\ell,i} \prod_{i=2}^n \omega_{i-1,i}^{k,\ell} \quad (4)$$

Then, with the degeneracy $G_C = \prod_{i=1}^n n_{\ell,i}$ of the subset C of conformations,

$$Z_S = \sum_{(C)} G_C \exp(-\beta E_S^C) = \sum_{(C)} (n_{h,1} \alpha_{h,1} \prod_{i=2}^n n_{\ell,i} \alpha_{\ell,i} \omega_{i-1,i}^{k,\ell}) \quad (5)$$

$$\text{This relation can be written as : } Z_S = U^* \left(\prod_{i=2}^n U_i \right) J \quad (6)$$

with the matrix $U_i = [u_{k,\ell,i}] = [n_{\ell} \alpha_{\ell,i} \omega_{i-1,i}^{k,\ell}]$, the row vector $U^* = [u_{h,1}] = [n_h \alpha_{h,1}]$ for $i = 1$ and the unit column vector J.

Monte Carlo samples of molecular conformations are obtained using these statistical weight matrices (ref. 3). The a priori probability that the unit (i) is in the map region or state ℓ can be computed using the relation (ref. 1) :

$$R_{\ell,i} = Z_S^{-1} U^* \left(\prod_{h=2}^{i-1} U_h \right) U'_{\ell,i} \left(\prod_{h=i+1}^n U_h \right) J \quad (7)$$

where the matrix $U'_{\ell,i}$ is deduced from U_i by replacing all elements by zero except those of column ℓ .

For a chain molecule of n units subject only to nearest neighbor interactions, the energy of a conformation taken in the subset C can be written as (ref. 4) :

$$E_S^C = E_1^h + \sum_{i=2}^n \epsilon_i^{k,\ell}$$

$$\text{with } E_1^h = -\beta^{-1} \ln \alpha_{h,1} \text{ and } \epsilon_i^{k,\ell} = -\beta^{-1} \ln (\alpha_{\ell,i} \omega_{i-1,i}^{k,\ell})$$

The average of E_S over the complete set of chain configurations is thus given by :

$$\langle E_S \rangle = \langle E_1 \rangle + \sum_{i=2}^n \langle \epsilon_i \rangle \quad (8)$$

$$\text{with } \langle \epsilon_i \rangle = Z_S^{-1} U^* \left(\prod_{h=2}^{i-1} U_h \right) U_i^+ \left(\prod_{h=i+1}^n U_h \right) J \quad (9)$$

$$\text{and } \langle E_1 \rangle = Z_S^{-1} U^{*+} \left(\prod_{h=2}^n U_h \right) J \quad (10)$$

where the element (k, ℓ) of the matrix U_i^+ is equal to $\epsilon_i^{k, \ell} u_{k, \ell; i}$ and the element h of the row vector U^{*+} is $E_1^h u_{h, 1}$.

The configurational free energy F_S and entropy S_S of the chain molecule can then be obtained from the following relationships :

$$F_S = - \beta^{-1} \ln Z_S = \langle E_S \rangle - T S_S \quad (11)$$

Chain models with short and long-range interactions

The total conformational energy E_T of a chain molecule is the addition of short-range and long-range interaction energies. The partition function of the chain is then given by :

$$Z_T = \sum_h \exp(-\beta \{E_S^h + E_L^h\}) \text{ while } Z_S = \sum_h \exp(-\beta E_S^h) \quad (12)$$

The summation is carried out over all the states h of the molecular system. The statistical average of a function f of the conformation should, in principle, be calculated with the relation :

$$\langle f \rangle = \sum_h f_h \exp(-\beta E_T^h) / Z_T \quad (13)$$

But, such quantities can only be evaluated by molecular dynamics methods or Monte Carlo sampling. In this last case, the calculation procedure we propose (ref. 3) consists of choosing chain conformations proportionally to statistical weights obtained from a chain model including only short-range interactions and then to keep or reject these conformations according to an importance sampling using the complementary part of the total energy coming exclusively from long-range interactions.

Note that chain models with only first neighbor interdependent units can be taken as references for the evaluation of the effects of long-range energy on chain configurations. Therefore, from relations (12) we get :

$$Z_T / Z_S = \langle \exp(-\beta E_L) \rangle_S \quad (14)$$

Angle brackets indicate an average over the chain configurations of the reference system (energy E_S). The free energy difference between the system ($E_S + E_L$) and the reference system is thus given by :

$$\Delta F = -\beta^{-1} \ln \langle \exp(-\beta E_L) \rangle_S \quad (15)$$

To improve the computation method we used a one-half umbrella sampling algorithm (refs. 6,7) for Monte Carlo calculations performed as described previously (ref. 3). Variations of the chain configurational entropy due to the introduction of E_L are estimated by using the average value of E_L obtained from an ordinary Monte Carlo sampling (ref. 8). Moreover, a series of N intermediate systems can be introduced (ref. 4) so that successive systems do present a good overlap of their configurational spaces ; we can then write :

$$\Delta F = \langle E_L \rangle - T\Delta S = -\beta^{-1} \sum_{j=0}^{N-1} \ln \langle \exp(-\beta E_L/N) \rangle_{S+(j/N)L} \quad (16)$$

$$\text{with } \langle \exp(-\beta E_L/N) \rangle_{S+(j/N)L} = Q(j+1)/Q(j) \quad (17)$$

$$\text{and } Q(j) = \sum_h \exp(-\beta \{E_S^h + (j/N)E_L^h\})$$

Averages corresponding to the different values of j are computed using a Monte Carlo procedure (ref. 3) in which E_S intervenes by means of matrices U and the fractions of E_L by Boltzmann factors. According to relation (13), $\langle E_L \rangle$ is obtained using this procedure.

RESULTS AND DISCUSSION

Calculation methods presented above were applied (with $T = 300^\circ\text{K}$) on simplified models of two peptidic hormones (refs. 9, 10) ; enkephalin : TYR-GLY-GLY-PHE-LEU (or MET) and β -casomorphin : TYR-PRO-PHE-PRO-GLY. Two different Monte Carlo calculations were performed. One of these procedures is based on the use of matrices U according to the method already described previously (ref. 3). The other Monte Carlo sampling method used is an ordinary one in which short-range interaction energy is calculated at each step of the procedure. The corresponding results thus obtained are indicated in Table 1.

The relative number of times a map region is affected to a unit of the chain gives the probability of occupancy of this region by the unit. Results obtained in this way can be compared to values calculated according to relation (7). We thus found that the study of region occupancies is a fine level of conformational analysis as it shows that the position of the relative minima in which the chains spend much time are actually different (ref. 3).

The determination of values of Z_S permits to get an estimation of F_S and,

TABLE 1

Values (Kcal/mole) of the free energy, energy and entropy terms.

	D	ΔF	$\langle E_L \rangle$	T ΔS
ENKEPHALIN	3.5	-30.2	-39.0	-8.8
	17.0	- 6.1	- 9.9	-3.8
	35.0	- 3.5	- 6.4	-2.9
	80.0	- 2.4	- 4.4	-2.0
β -CASOMORPHIN	3.5	-20.0	-25.6	-5.6

REFERENCE CHAIN MODELS

ENKEPHALIN F_S : -20.9 ; $\langle E_S \rangle$: -10.0 (U) ; TS_S : 10.9
- 9.6 (M)

β -CASOMORPHIN F_S : -26.9 ; $\langle E_S \rangle$: -19.6 (U) ; TS_S : 7.3
-18.7 (M)

D : dielectric constant.

U : calculations with matrices U.

M : ordinary Monte Carlo.

with $\langle E_S \rangle$ already known, the corresponding value of the entropic term TS_S can be calculated.

Results of calculations show clearly that long-range interactions do represent the major contributions to the conformational energy. Their effect on the chain configuration is mainly to lead zwitterionic molecules toward folded forms. Indeed, we have observed that when interactions between the charged ends of molecules are screened, the probabilities for a unit to be in a map region, calculated from Monte Carlo sampling with a completely interacting chain model, are very much similar to values obtained by using chain models with nearest neighbor interdependent units.

The importance of E_L being essentially due to electrostatic interactions, sampling calculations were performed with different values given to the dielectric constant D. Results thus obtained are presented in Table 1. The different values obtained for T ΔS , depending on D, show that the loss of flexibility of chain molecules, folded by long-range interactions, can be well estimated by

the variations of the configurational free energy and entropy.

Note that the definition of E_L (and even E_S) may include a potential of mean force in order to simulate in a realistic way solvent effects on molecular conformation. Moreover, the calculation methods presented are completely general and can be applied on very different kinds of chain models.

REFERENCES

1. P.J. Flory, Statistical Mechanics of Chain Molecules, Interscience, New York (1969).
2. M. Volkenstein, Configurational Statistics of Polymeric Chain, Interscience, New York, (1963).
3. S. Prémilat, J. Chem. Phys. 88, 3385 (1988).
4. S. Prémilat, J. Chem. Phys. 90, 5128 (1989).
5. F.A. Momany, R.F. McGuire, A.W. Burgess and H.A. Scheraga, J. Phys. Chem. 79, 2361 (1975).
6. G.M. Torrie and J.P. Valleau, J. Comp. Phys. 23, 187 (1977) ; Chem. Phys. Lett. 28, 578 (1974).
7. C.Y. Lee and H.L. Scott, J. Chem. Phys. 73, 4591 (1980).
8. S. Prémilat and B. Maigret, J. Chem. Phys. 66, 3418 (1977).
9. P.W. Schiller, The Peptides, Vol. 6, Academic, New York, (1985).
10. V. Brantl, H. Teschemacher, J. Blasig, A. Henschen and F. Lotspeich, Life Sci. 28, 1903 (1981).

DISCUSSION

WIPFF - 1. What size of chains are you focusing on ? From the examples shown, it seems to deal mainly with short polypeptides.

2. I am somewhat puzzled about the relevance of these simulation for conformational free energies in solution. How would you treat the effective interaction between charged groups and the hydrophobic effects on folding ?

PREMILAT - 1. Now, I am just using short polypeptide chains, mainly in order to test the calculation methods. But all the algorithms can be applied on longer chains (not too long because M.C. is computer time consuming). The number of regions defined in the (ϕ, ψ) maps is also an important factor because it defines the order of the matrices of averaged statistical weights.

2. I think that in these kind of simulations, the solvent effect must be taken into account by some potential of mean force rather than by a molecular model of water.

PETTIT - Is there a self consistency or feed back between the short range correlations and the long range correlations ; that is, do they modify each other ?

PREMILAT - The energy of the molecular chain is exprimed by an addition of terms, so if any correlation between short-range and long-range interactions exists, it can be analyzed with the M.C. samples of conformations. One can determine the probabilities of occupancy of the different map regions by every chain unit and then see if they are modified when long-range interactions are included in calculations. This will give some information on the relative importances of short and long-range energies and on the correlations between them.

EVALUATION OF ELECTROSTATIC PROPERTIES AT ENZYME ACTIVE SITES

G. DIVE, J. LAMOTTE-BRASSEUR and D. DEHARENG

Service de Microbiologie, Université de Liège, Institut de Chimie, B6,
B-4000 Sart Tilman (Liège 1), Belgium

SUMMARY

This paper is devoted to the electrostatic potential characteristics of a model enzyme (α -chymotrypsin) active site and to fast derivations, within an approximate quantum chemistry scheme, of the electrostatic interaction energy as an efficient tool to predict stable conformations of the partners in an enzymatic reaction.

INTRODUCTION

Life related sciences have always been very appealing investigation domain for theoretical study but were very difficult to investigate with sufficient reliability due to the dimensions of the systems involved. Nowadays, some of these problems are becoming easier to handle because of the development of computer technology, particularly in one of the most crucial challenge proposed by the study of the interaction between two partners as an enzyme model active site and its ligand.

A very physically appealing way to study the interaction between several partners is to split the interaction energy into well-defined terms as in the Morokuma's decomposition scheme (ref. 1). That the electrostatic forces are most often the leading strength which determines the relative orientation of two or more partners of a complex is already well-known (ref. 2) and in this respect, the electrostatic potential (ref. 3) has been interpreted as a reactivity index (ref. 4).

In this work, we wanted to emphasize the characteristics and the usefulness of two electrostatic quantities or properties : the electrostatic potential on one hand and the electrostatic interaction energy on the other.

THE ELECTROSTATIC POTENTIAL

Given a charge distribution ρ in a medium characterized by a polarizability \vec{P} and a dielectric constant ϵ , it results an electrostatic field \vec{E} and an electrostatic potential V altogether related by :

$$\vec{\nabla} \cdot \vec{E} = - \vec{\nabla} \cdot (\vec{\nabla} V) = 4 \pi (\rho - \vec{\nabla} \cdot \vec{P}) = 4 \pi \frac{\rho}{\epsilon} \quad (1)$$

The electrostatic potential V (E.P.) is more widely used than \vec{E} as this scalar quantity represents the interaction energy between a molecule with the charge distribution ρ and a bare unitary positive point charge. Thus, E.P. reveals in a straightforward way the nucleophilic regions of the system and their relative magnitude. As such, it has been considered as a reactivity index of a molecule (ref. 4).

In the quantum chemistry framework, ρ is expressed in terms of molecular orbitals obtained in a self-consistent way, *i.e.*, which already take into account such effects as the polarization of each charged particle on its surrounding partners, as well as exchange repulsion of the electrons, charge transfer, and so on. Hence, the derivation of V from a quantal ρ must not involve any other polarizability and in eqn. (1), $\vec{\nabla} \cdot \vec{P}$ is set to zero and ϵ to 1. Then, V is written :

$$V(r) = \sum_{\mu} \sum_{\nu} D_{\mu\nu} \int dr' \chi_{\mu}(r') \frac{1}{|\vec{r}' - \vec{r}|} \chi_{\nu}(r') + \sum_{\alpha} Z_{\alpha} / |\vec{R}_{\alpha} - \vec{r}| \quad (2)$$

where $D_{\mu\nu}$ are the density matrix elements in the basis of the $\chi_{\mu}(r')$ atomic orbitals, Z_{α} is the nuclear charge of α .

The adequacy of the use of a dielectric function as a screening function in other contexts is discussed in ref. 5. In this work, all the calculations have been carried out within the consistent quantum chemistry framework, according to eqn. (2).

Polarization corrective term

Nevertheless, there exists another possible correction not to be confused with this previous dielectric effect : it is the polarization response of the whole molecule coming from the charge distribution of the second partner it interacts with. In the case of a bare proton, the correction V_{PL} to V can be evaluated by the perturbation theory method and can be approximated, in a SCF treatment by (ref. 6) :

$$V_{PL}(r) = \sum_i^{NO} \sum_a^{NV} (\epsilon_i - \epsilon_a)^{-1} \left[\sum_{\mu} \sum_{\nu} c_{\mu i} c_{\nu a} \int dr' \chi_{\mu}(r') \frac{1}{|\vec{r} - \vec{r}'|} \chi_{\nu}(r') \right]^2 \quad (3)$$

where NO, NV are the number of occupied and virtual M.O. respectively, ϵ_i and $c_{\mu i}$ are the M.O. eigenvalues and eigenvectors. This correction V_{PL} may be of some importance in specific problems like the determination of proton affinities (ref. 7). It is highly sensitive to the inclusion of polarization atomic orbitals (ref. 8), particularly with small basis sets.

However, the effect of the polarization given by a unitary point charge is effectively much higher than that generated by a real system. Hence, this correction is not expected to lead to inversion of the electrostatic behavior of neutral interacting species.

Calculation of the electrostatic potential at α -chymotrypsin active site

As example of the E.P. usefulness in enzyme ligand interaction study, several active site models of α -chymotrypsin have been investigated. The potential features were apprehended *via* 3D maps at -10 kcal/mole (ref. 5). Two distinct ways of solving eqn. (2) have been considered, the first without approximation but using $D_{\mu\nu}$ obtained after deorthogonalization of CNDO/2 M.O., the second with the approximation hereafter called the γ approximation (ref. 9) :

$$V(r) = - \sum_{\alpha} q_{\alpha} \gamma_{\alpha H} + \sum_{\alpha} Z_{\alpha} / |\vec{r} - \vec{R}_{\alpha}| \quad (4)$$

where q_{α} are the charges obtained after a Mulliken population analysis, and $\gamma_{\alpha H}$ are the bielectronic repulsion CNDO/2 (ref. 10) integrals between s orbitals on α and on H^+ at point r . The contours at -10 kcal/mole obtained at the two levels of V in the case of the smaller model active site of α -chymotrypsin (91 atoms), are very similar and thus, the potential maps were calculated within the much faster approximation for the bigger active site models (207 and 216 atoms).

One of the most striking features of the potential is the clearcut cooperativity of the carbonyl groups of the protein backbone to design an impressive negative region on one side of the molecule by merging their respective potential wells. As a counterpart to these favourable folded conformations generating such a large negative potential cloud, no negative well at this level is drawn around β strand substructures (Fig. 1).

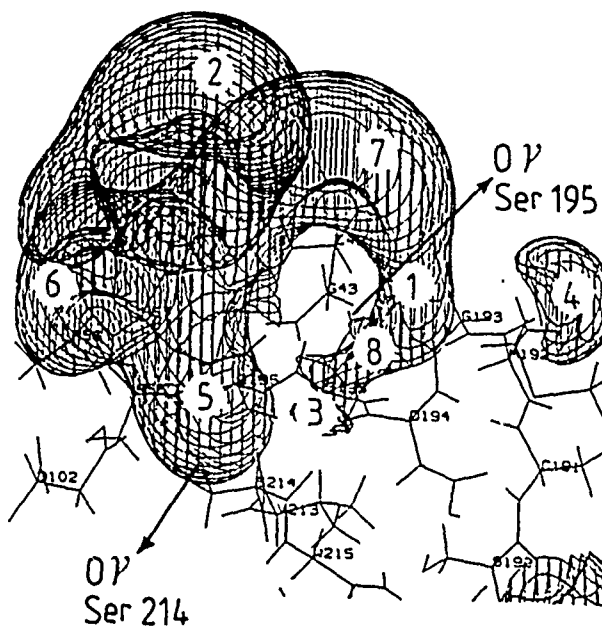


Fig. 1. E.P. contour map at -10 kcal/mole for a model active site of α -chymotrypsin (216 atoms). The subregions 1 to 8 are discussed in ref. 5.

This complementarity can be related to the high dipole moment value observed in each model and emphasizes the directionality of the driving force involved in the formation of a Michaelis complex.

The electrostatic interactions are the leading strength of the first step of the enzymic reaction and thus the electrostatic potential constitutes a very useful qualitative tool in the understanding of the beginning of the process. However, in order to go further and quantify the electrostatic interaction between the two partners, their electrostatic interaction energy has to be calculated.

THE ELECTROSTATIC INTERACTION ENERGY

In the case of molecular systems, the electrostatic interaction energy writes :

$$E_{\text{int}}^{\text{ES}} = \iint d\mathbf{r} d\mathbf{r}' \frac{\rho_A^{\text{el}}(\mathbf{r}) \rho_B^{\text{el}}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} - \sum_{\alpha} Z_{\alpha} \int d\mathbf{r}' \frac{\rho_B^{\text{el}}(\mathbf{r}')}{|\mathbf{R}_{\alpha} - \mathbf{r}'|} - \sum_{\beta} Z_{\beta} \int d\mathbf{r} \frac{\rho_A^{\text{el}}(\mathbf{r})}{|\mathbf{R}_{\beta} - \mathbf{r}|} + \sum_{\alpha} \sum_{\beta} Z_{\alpha} Z_{\beta} / |\mathbf{R}_{\alpha} - \mathbf{R}_{\beta}| \quad (5)$$

where ρ^{el} is the electronic density function of partner A or B. In order to obtain numerically good results (ref. 11), the four terms of eqn. (5) must be grouped :

$$E_{\text{int}}^{\text{ES}} = \int d\mathbf{r} V_A(\mathbf{r}) \rho_B^{\text{el}}(\mathbf{r}) + \sum_{\beta} Z_{\beta} V_A(\mathbf{R}_{\beta}) \quad (6)$$

where $V_A(\mathbf{r})$ is the electrostatic potential of partner A. It is obvious that the second term of eqn. (6) is very easily calculated once V_A is known. The first term can be determined either analytically or by a 3D numerical integration. The analytic electrostatic energy is computed at the STO-3G W1S level (without the 1s orbitals on the heavy atoms) by our modified link 604 of GAUSS82 implemented on a Data General MV7800 computer (0.8 Mips).

A. Numerical derivation

With a well-conditioned stepsize of 0.25 Å of the 3D grid, the calculation of one value of $E_{\text{int}}^{\text{ES}}$ lasts 25 s compared to 103 s for an analytical STO-3G W1S derivation for the complex between the dyad water-imidazole (partner A) and the methanol (partner B) (Fig. 2). Hence, the numerical procedure constitutes an appreciable gain of computer time as the determination of the most stable conformations requires many calculations.

The question remains as to which level of approximation V_A and ρ_B^{el} are determined with. Two levels are presented here.

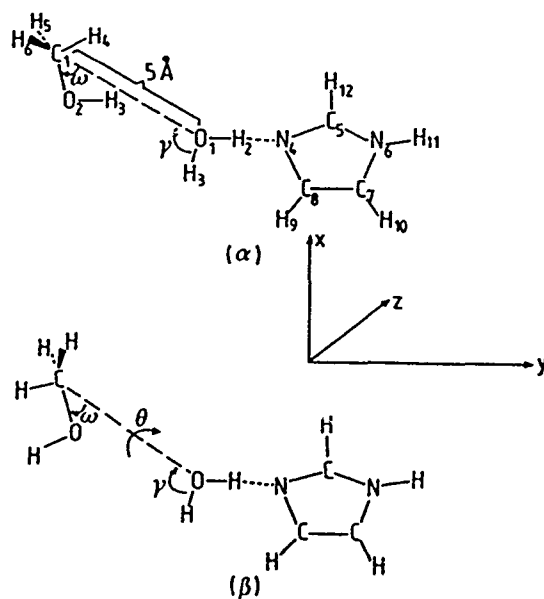


Fig. 2. Definition of the rotational coordinates for two configurations (α, β) of the complex methanol + the dyad water-imidazole.

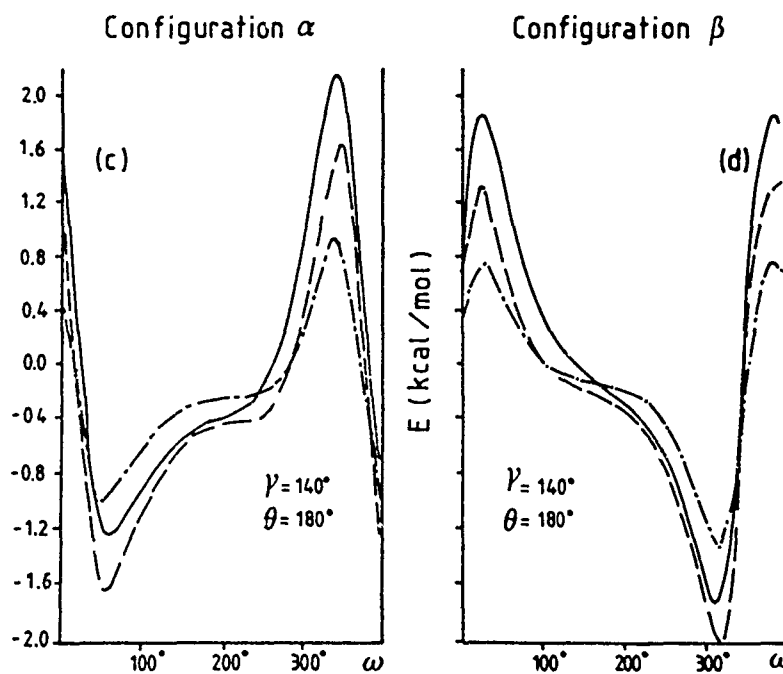


Fig. 3. Electrostatic interaction energy between the methanol and the dyad water-imidazole for the two configurations (α, β) . Full line : analytical STO-3G WIS result. Dashed line : numerical result at level (a). Dash-dotted line : numerical result at level (b).

Level (a) : the grids are determined with deorthogonalized CNDO/2 density matrix elements $D_{\mu\nu}$ and eqn. (2) is used for calculating V_A .

Level (b) : ρ_B^{el} is obtained at level (a) but V_A is calculated according eqn. (4), within the γ approximation.

The rotational coordinates are defined in Fig. 2. As it can be seen in the particular case of the coordinate ω , the two levels provide very good results compared to the analytical reference (Fig. 3).

As a conclusion, the numerical procedure runs 4 times faster than the analytical one for the studied complex and gives very satisfactory results concerning relative orientation of the two partners at a given intersystem distance. Thus, it constitutes a powerful tool for predicting stable conformations of complexes in a region where the electrostatic approximation is valid, at internuclear distances between the partners greater than 2.5 Å.

B. Approximate analytical derivation

In the same way, we wanted to study whether it was possible to obtain good results with an analytical procedure. Furthermore, in order to remain within reasonable CPU times to deal with big systems, only approximate methods were to be considered. It was chosen to work within the CNDO/2 framework (ref. 10), in which the first three terms of eqn. (5) write :

$$E_{int,1}^{ES} = \sum_{\alpha \in A} \sum_{\beta \in B} \gamma_{\alpha\beta} \left(\sum_{\mu \in \alpha} \sum_{\nu \in \beta} D_{\mu\mu}^{\alpha} D_{\nu\nu}^{\beta} \right) \quad (7)$$

$$E_{int,2}^{ES} = - \sum_{\alpha} Z_{\alpha} \sum_{\beta} \gamma_{\alpha\beta} \sum_{\mu \in \beta} D_{\mu\mu}^{\beta} \quad (8)$$

$$E_{int,3}^{ES} = - \sum_{\beta} Z_{\beta} \sum_{\alpha} \gamma_{\alpha\beta} \sum_{\nu \in \alpha} D_{\nu\nu}^{\alpha} \quad (9)$$

The terms $\left(\sum_{\mu \in \alpha} D_{\mu\mu}^{\alpha} \right)$ can be considered as the electronic charges localized on the nuclei α belonging to partner A. They can be evaluated either at CNDO/2 level or derived from the Mulliken population analysis applied on the MO coefficients (hereafter called CNDO/D approximation). A comparison between the two kinds of derivation has been analyzed in the study of the complex dyad water-imidazole and the methanol. Though the similar curves thus obtained vary much more smoothly and present lower absolute values than the analytical STO-3G WIS ones, they behave quite correctly and the largest difference in the position of the minima is of only 10° for the considered rotational coordinates. For every curve, the CNDO/D results were better than the simple CNDO/2 ones and in the following, all the calculations were done at the CNDO/D level.

At this point of the discussion, the adequacy of using an approximately calculated (numerically or analytically) E_{int}^{ES} to find out the most stable relative orientation of two partners has been investigated. The influence of their

internuclear distances was negligible because none were lower than 2.5 \AA . What goes on when, during a rotation or during a direct approach, one or more inter-system distance gets too short? As can be seen from Fig. 4, the electrostatic

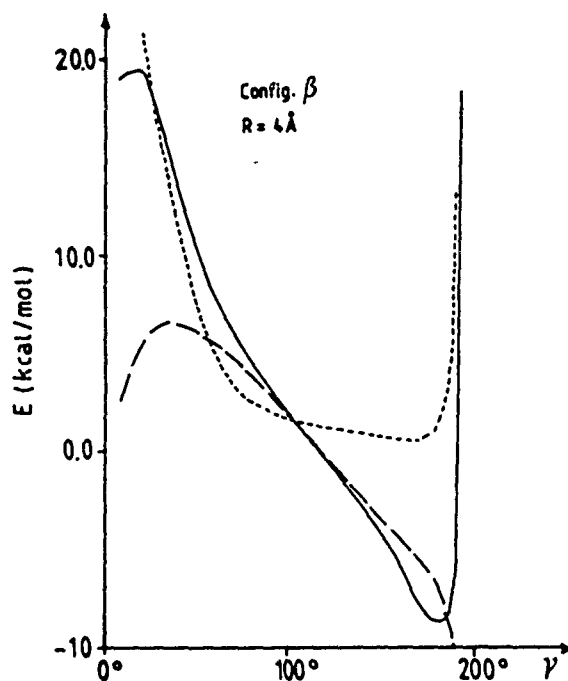


Fig. 4. Electrostatic $E_{\text{int}}^{\text{ES}}$ and total E_{int} interaction energy between the methanol and the dyad water-imidazole in the configuration β . Full line: E_{int} at the 3-21G level. Dashed line: $E_{\text{int}}^{\text{ES}}$ at the 3-21G level. Dotted line: $E_{\text{int}}^{\text{ES}}$ at the CNDO/D level.

energy derived with the 3-21G basis set (this is also true for other basis sets) from a Morokuma's decomposition has a completely different shape than the total interaction energy though they have similar values over a large part of the variation domain of the coordinate γ . It can be noted that the differences in E_{int} energy between ST0-3G and 3-21G results are significant. However, when γ is greater than 180° , the distance between the oxygen of methanol and one of the hydrogen of imidazole becomes shorter than 2.5 \AA and other components of E_{int} are then important, particularly the exchange repulsion term: E_{int} reaches a minimum which does not exist at this value of γ for $E_{\text{int}}^{\text{ES}}$. Surprisingly, the electrostatic CNDO/D energy does exhibit a minimum at this γ value. That $E_{\text{int}}^{\text{ES}}$ (CNDO/D) presents a minimum long before the *ab initio* correct one and even before the total interaction SCF energy has been pointed out by Sokalski *et al.* (ref. 12) for the system $(\text{H}_2\text{O})_2$ and by ourselves for the system $\text{H}_2\text{O} + \text{formamide}$ (ref. 13). This misbehavior is due to the fact that the fourth term of eqn. (5) i-

calculated correctly while the attractive resultant, sum of the first three terms, is obtained within the CNDO/D approximation. This drawback of the method provides however a good equilibrium value of γ in the case of the larger system methanol + (water-imidazole). This led us to study what happened when many inter-nuclear distances were involved in the relative orientation of two partners, *i.e.* when big complexes were considered. This type of study was made on the model active site of α -chymotrypsin made up of 22 amino acids (see Table 1 of ref. 5) (262 atoms) and the substrate N-acetyl-L-tryptophanamide (33 atoms). The values of E_{int}^{ES} (CNDO/D) were compared with the total interaction energy obtained at the CNDO/2 level. For comparison, one E_{int}^{ES} (CNDO/D) value is obtained after *ca.* 45 s on a VAX 11/780 whereas one E_{int} (CNDO/2) calculation lasts *ca.* 5800 s on a FPS264 attached processor (38 mflops). This gain of time is obviously appreciable.

One local minimum was found for E_{int}^{ES} (CNDO/D) that coincided with one for E_{int} (CNDO/2) and the energy differences between this point and two neighbors were very similar for the two functions. However, further investigation is needed to confirm this similarity in E_{int}^{ES} (CNDO/D) and E_{int} (CNDO/2) behavior.

ACKNOWLEDGEMENTS

This work was supported in part by the *Fonds de la Recherche Scientifique Médicale* (contract n°3.4537.88), an *Action Concertée* with the Belgian Government (convention 86/91-90), a convention with the *Région wallonne* (C2/C16/Conv.246/20428), the *Fonds de Recherche de la Faculté de Médecine ULg* and a contract with the *EEC* (BAP-0197-B). GD is *chercheur qualifié* of the *FNRS*, Brussels.

REFERENCES

- 1 K. Kitaura and K. Morokuma, A new energy decomposition scheme for molecular interactions within the Hartree-Fock approximation, *Int. J. Quantum Chem.*, X (1976) 325-340.
- 2 R. Bonaccorsi, C. Ghio and J. Tomasi, On a semiclassical interpretation of inter and intramolecular interactions, *Int. J. Quantum Chem.*, XXVI (1984) 637-686.
- 3 E. Scrocco and J. Tomasi, The electrostatic potential as a tool for the interpretation of molecular properties, *Top. Curr. Chem.*, 42 (1973) 95-170.
- 4 J.S. Murray and P. Politzer, Electrostatic potentials of amine nitrogens as a measure of the total electron-attracting tendencies of substituents, *Chem. Phys. Lett.*, 152 (1988) 364-370.
- 5 J. Lamotte-Brasseur, G. Dive, D. Dehareng and J.M. Guhysen, Electrostatic potential maps at the quantum chemistry level of the active sites of the serine peptidases. α -Chymotrypsin and subtilisin, *J. Theor. Biol.*, submitted.
- 6 M. Miller-Francé, MEPHISTO : a program for the calculation of molecular electrostatic potentials including polarization corrections, QCPE n°490, Department of Chemistry, Princeton University, Princeton, N.J. 08544.
- 7 H. Umeyama and K. Morokuma, Origin of alkyl substituent effect in the proton affinity of amines, alcohols and ethers, *J. Amer. Chem. Soc.*, 98 (1976) 4400.
- 8 D. Dehareng, G. Dive and J.M. Guhysen. Polarization corrections to the electrostatic potential, in preparation.
- 9 C. Giessner-Prettre and A. Pullman, Molecular electrostatic potentials : comparison of *ab initio* and CNDO results, *Theor. Chim. Acta*, 25 (1972) 83-88.

- 10 J.A. Pople and D.L. Beveridge, Approximate Molecular Orbital Theory, Mc Graw-Hill, New York, 1970.
- 11 D. Dehareng, G. Dive, J. Lamotte-Brasseur and J.M. Chuysen, Numerical computation of the electrostatic interaction energy between methanol and the dyad water-imidazole, Theor. Chim. Acta, 76 (1989) 85-94.
- 12 W.A. Sokalski and H. Chojnacki, Approximate exchange perturbation study of intermolecular interactions in molecular complexes, Int. J. Quantum Chem., XIII (1978) 679-692.
- 13 D. Dehareng, G. Dive, J. Lamotte-Brasseur and J.M. Chuysen, Analytical calculation of the electrostatic energy within the CNDO framework, in preparation.

DISCUSSION

SURCOUF - Did you analyse the incidence of the size of your model active site on the consistency of your results ?

DEHARENG - The results present electrostatic potential maps of the active sites of α -chymotrypsin and subtilisin. An extensive study was made in the case of α -chymotrypsin in which the progressive building of the model active site from 6 up to 19 amino acids clearly indicated the general tendency of the negative cloud of the electrostatic potential to remain localized on one part of the active site. Though its detailed shape is a little modified from one model to the other, its main characteristics remained quite similar. Thus afterwards, the choice of the number of amino acids taken into account in subtilisin was made on the same footing as for the biggest active site model of α -chymotrypsin.

INHIBITOR AND SUBSTRATE BINDING IN DIHYDROFOLATE REDUCTASE - A THEORETICAL APPROACH*

T. ŠOLMAJER^{1,2} and M. HODOŠČEK¹

¹ Boris Kidric Institute of Chemistry, Hajdrihova 19, 61000 Ljubljana, Yugoslavia

² Lek- Pharmaceutical and Chemical Industry, Celovska 135, 61000 Ljubljana, Yugoslavia

SUMMARY

A fundamental problem in chemistry and biochemistry of proteins is understanding the role of a single amino acid residue in determining the molecular properties of the protein. Structural information from refined X-ray structures of several species of the enzyme dihydrofolate reductase with the bound inhibitor methotrexate have been reported and this provides a firm experimental basis for theoretical investigation of changes in inhibitor binding brought about by a single amino acid substitution of the parent enzyme. We have applied a novel methodology consisting of combined use of *ab initio* and thermodynamic integration methods implemented with molecular dynamics to determine the binding free energies of methotrexate and the natural substrate dihydrofolate theoretically. "Hydrogen bond adapted" parametrization of potential gives encouragingly good agreement with experimental free energy values.

INTRODUCTION

Dihydrofolate reductase (DHFR: tetrahydrofolate dehydrogenase E.c.1.5.1.3) is a key enzyme involved in biosynthetic pathways leading to the production of the purine and pyrimidine nucleotides. Specifically, it catalyzes the NADPH dependent reduction of 7,8-dihydrofolate (H₂F). This enzyme has been the subject of intensive investigation for over three decades(ref. 1). Such continuing interest has been prompted by its importance as biological target for a large class of drugs - the antifolates.

Methotrexate (MTX) is a potent inhibitor of the enzyme and still one of the most widely used antineoplastic agents. The principal difference between the inhibitor MTX and the natural substrate dihydrofolate (H₂F) is the 4-amino group on MTX which replaces the 4-oxo group on the H₂F. It appears from the extensive structure-activity relationship studies on antifolates(ref. 2) that the main binding characteristics necessary for high binding constant of the inhibitor ($K_D = 0.07$ nM) is the 2,4 - diamino pyrimidine ring. Already Baker(ref. 3) postulated that the increased basicity of the pteridine ring achieved by the 4-amino substituent permits strong interaction with acidic groups in the active site of the enzyme.

This interaction between DHFR and MTX has been studied well both experimentally (refs. 4-5) and theoretically (refs. 6-8). Significant breakthrough in the study of molecular basis for the tight binding was achieved by solving for the crystal structures of binary complex DHFR:MTX and ternary complexes (with NADPH coenzyme) to high resolution(refs. 9-10).

* Part of this work was presented at the International Symposium: Molecular Recognition - Its Role in Chemistry and Biochemistry, Sopron (Hungary), 24-27.08.1988

These structures revealed that the Asp 27 residue of the wild type DHFR forms a pair of strong hydrogen bonds with N1 and 2-amino groups of MTX (Fig. 1).

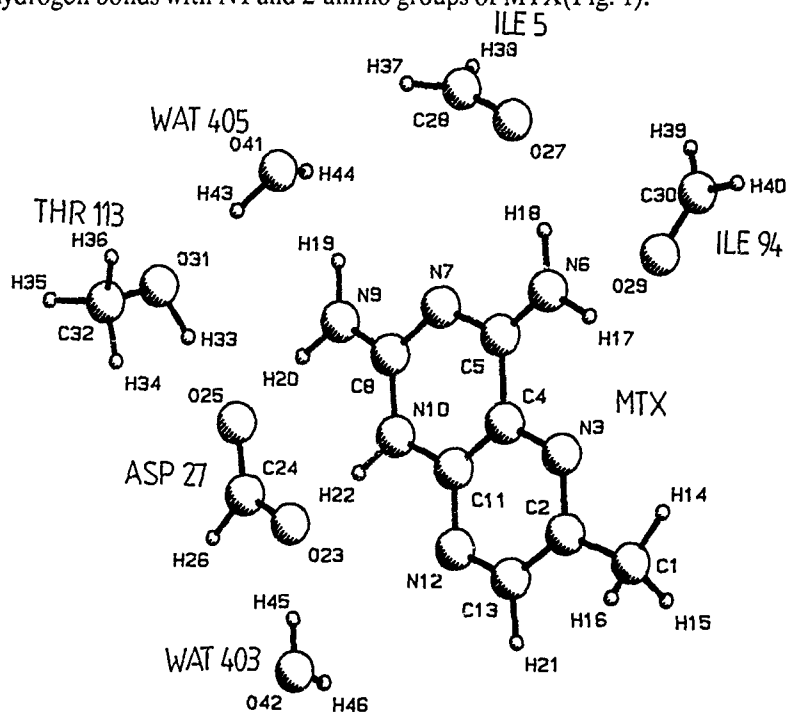


Fig. 1: Active site of wild type DHFR with bound MTX (X-ray geometry from ref. 9) used for ab initio calculations. MTX is modelled by 6-methyl,2,4- diamino pteridine. Similarly, amino acid residues Ile5,Asp27,Ile94,Thr113, Wat403 and Wat405 are modelled by taking their respective hydrogen bond forming parts into account.

The authors have also proposed that the substrate is flipped over when bound to the active site of DHFR with respect to the position of MTX. Recently, this tentative model was experimentally proven by the X-ray study of Winkler et al.(ref. 11).

Further important advances in understanding the mode of action of the DHFR enzyme came from the solved crystal structures of Asp27->Ser27 and Asp27->Asn27 mutants(ref. 12) obtained by site-directed mutagenesis. The geometry of the active site interactions of mutant enzymes did not change with respect to the wild type enzyme except for changes in hydrogen bond pattern due to the point amino acid residue mutations. In addition, K_D values were determined showing that while MTX binds about 27 times less strongly to the Asn27 DHFR ($K_D = 1.9$ nM), the Ser 27 mutant has a $K_D = 210$ nM (ref. 12). Another interesting aspect of binding to mutant enzymes was determined by the difference spectroscopy experiments. The state of protonation of inhibitor MTX has changed when going from wild type enzyme to mutants: MTX is protonated in the wild type enzyme, while when bound to either mutant it is not(ref. 13).

Thus, there exists a wealth of experimental data on this enzyme system which makes it a very suitable basis for theoretical studies of correlation of the structure and action on the molecular level. Such correlations require knowledge of the corresponding relation between structure and energy (ref. 14). Probably the most important factors in such a structure-activity

correlation are associated with electrostatic interactions. Electrostatic interactions which are involved in hydrogen bonding and salt bridges are important in general for maintaining the threedimensional structure of proteins. However, when these structural elements are present in the active site regions of enzymes they become even more important as key elements in binding of substrates and inhibitors(ref. 15).

In order to achieve a quantitative understanding of action of biological molecules one needs the ability to correlate electrostatic interactions with structural information. The difficulty of such a task is apparent from the following example: the electrostatic free energy of a point charge in a polar solvent is about - 80 kcal/mol as calculated by the well known Born formula. On the other hand, the experimentally determined difference in binding free energy of protonated and unprotonated methotrexate to the active site of the enzyme DHFR is only about 1.8 kcal/mol.

We show on a well defined system DHFR:MTX and DHFR:H₂F how the hydrogen bonds donated by specific residues of the protein or fixed water molecules control the inhibitor and substrate binding. The mechanism of charge stabilization in the active site of DHFR is studied by ab initio quantum mechanical method for the native enzyme and the two mutants Asp27->Ser27 and Asp27->Asn27 with bound MTX and H₂F, respectively. The calculated microscopic picture of changes in electron density (refs. 17-19) brought about by mutation is subsequently used in the statistical mechanical calculations of binding free energy(refs. 20-22). The latter has been successfully used for calculations of inhibitor binding and catalyses of substrates, however, with a restriction that the potential force field used should reasonably describe the salient features of the process energetics. Here we use the advantages of two powerful methods: ab initio treatment and molecular dynamics approach to calculate small, but highly significant changes in the enzyme active site.

METHODS

Thermodynamic cycles approach and basic theory of free energy calculations have been extensively described in literature (refs. 20-22 and references therein). Therefore we reiterate here only one key feature: we simulate only the enzyme -> mutant enzyme, and enzyme : ligand -> mutant enzyme : ligand systems and thus avoid the more inaccurate (at least in case of charged amino acid residues) solution simulation (ref. 7). Below we give a summary of computational details.

Ab initio calculations were performed with standard 3-21G basis set and using the Monstergauss package (ref. 23). Each system, describing the enzyme active site with a bound ligand consisted of 45-50 atoms described by approximately 250 basis functions. In order to test the approximation we have chosen, i.e. how well the active site model describes the actual situation in the enzyme, we have also incorporated the charge distribution of the enzyme and crystal water molecules directly into the ab initio Hamiltonian by using point charges from the molecular dynamics residue library (ref. 24). Approximately 2100 point charges were included in the ab initio Hamiltonian. Electron density maps were computed on a grid of approximately 1000 points.

Molecular dynamics simulations were done (ref. 24) on systems of about 1550 enzyme atoms using the united atom approximation and about 200 water molecules. Wall region of radius 18 Å centered around N1 atom of MTX and N3 atom of H₂F, respectively, was employed with the nonbond cutoff distance of 9 Å. Atoms outside this region were kept frozen. Energy minimization was followed by a short MD trajectory to allow the system to reach equilibrium at 300 K and then the amino acid in question was subject to mutation for the native enzyme, and enzyme:inhibitor (substrate) complex, respectively. Geometry of the enzyme:substrate complex (Fig. 2) has been modelled after the X-ray data of Winkler et al. (ref. 11). The molecular dynamics simulation implemented with the thermodynamic integration technique was performed using isothermal- isobaric conditions.

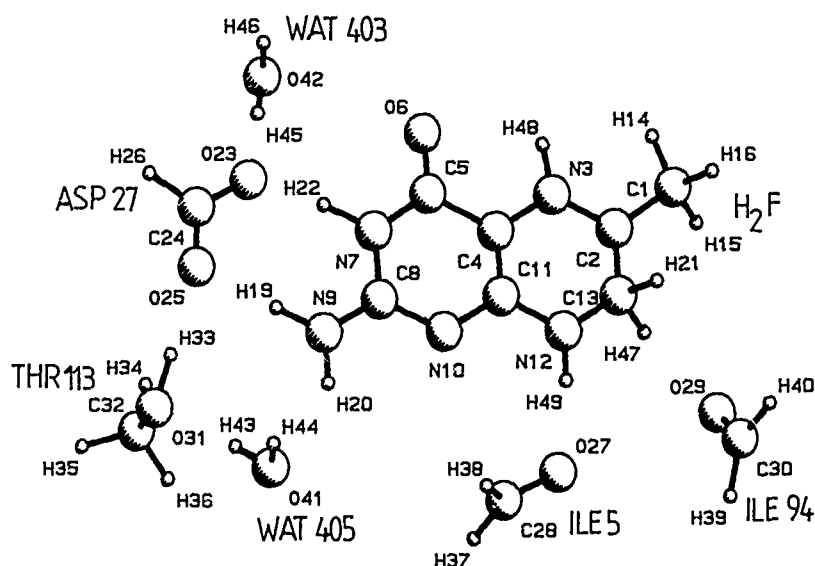


Fig.2 Active site of wild-type DHFR with bound substrate (6-methyl,7,8-dihydropterin) Same as in Fig.1.

Weak coupling to an external temperature bath of 300 K with a coupling constant of 0.4 ps and to an external pressure bath of 1.0 bar with a coupling time constant of 0.5 ps was used to maintain temperature and pressure in all systems. The leap frog algorithm was used with velocity and coordinate rescaling to accomplish the weak coupling. The model potential for the water intermolecular interactions used was the simple point charge (SPC) developed by Berendsen et al (ref. 25). It has been shown to describe well the equilibrium properties and free energies of hydration in the molecular dynamics simulation of ionic hydration (ref. 26). The intramolecular degrees of freedom were treated as separable constant using the SHAKE coordinate resetting procedure. The time step used was 0.002 ps. Free energy calculations were done "forward" and "backward" in 20 ps. This proved to be sufficient because of the linearity of the free energy change dependence on time (ref. 27). We used the nonlinear parametrization proposed by Cross (ref. 25) for the Van der Waals potential term and ab initio determined charge distribution parameters.

RESULTS AND DISCUSSION

The results of proton potential calculations with the ab initio method are given in Tables 1 and 2 for the MTX and H₂F, respectively. The plot of the potential for the MTX complex against the N1-H distance is given in Fig.3. Potential energy of the proton in the hydrogen bond of the wild type enzyme shows that the salt bridge structure is energetically about 19 kcal/mol more favourable than neutral hydrogen bond with proton at Asp27 oxygen. This figure is reduced to 17.5 kcal/mol when all protein and crystal water charges are taken into account.

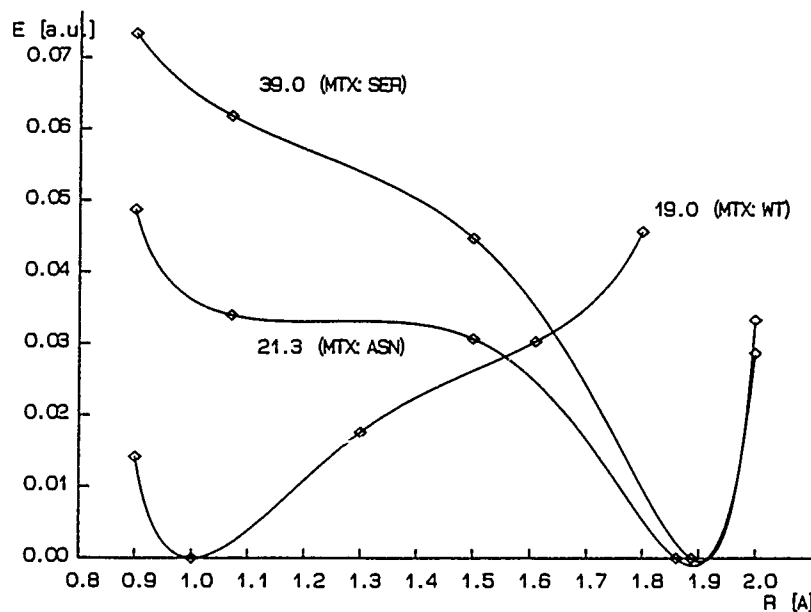


Fig.3: Proton potential for the proton (H22) in the hydrogen bond between N10 atom of pteridine ring and respective side chain atom of residue 27. (a) MTX:WT hydrogen bond is to OD2 atom of Asp27 (b) MTX:ASN hydrogen bond is to ND2 of mutant Asn27 (c) MTX:SER hydrogen bond is to OG of mutant Ser27. R is the distance from N10 atom of MTX in Å.

TABLE 1

Energetics of proton transfer in the hydrogen bond N1 (MTX) - residue 27 (DHFR). Energy values are relative to the minimum of proton potential function. Units: kcal/mol

	R (N1-H) [Å]	E [kcal/mol]
Asp 27	1.02	0.
	1.61	19.0
Asp 27	1.02	0.
with p.p.c.	1.61	17.5
Asn 27	1.02	21.3
	1.86	0.
Ser 27	1.02	27.9
	1.87	0.

Our result can be compared with recently published calculation by Singh (ref. 7). His simple model included MTX with Asp27 residue only and the potential energy minimum was found to be on the neutral Asp side. The preference for a salt bridge in the guanidinium moiety bound to carboxylate was found previously for a series of substituted guanidinium isosteres (refs. 27-28) mimicking the H2-receptor antagonists, however in this case the energy difference between the two potential minima was significantly smaller (about 6 kcal/mol). Clearly, other amino acid residues of the active site forming hydrogen bond network around the inhibitor are of importance. This is further corroborated by the fact that the binding of MTX to Thr113 -> Val113 mutant is about 30 times less strong than is the case with the wild type enzyme (ref. 31). In Table 2 the proton potential for the substrate in the three complexes is given (ref. 27). In the wild-type DHFR the potential minimum is much more shallow and, in fact, (not shown here for the lack of space) also of a double minimum type.

TABLE 2

Energetics of proton transfer in the hydrogen bond N3 (HF) - residue 27 (DHFR). Energy values are relative to the minimum of proton potential function. Units: kcal/mol

	R (N3-H) [Å]	E [kcal/mol]
Asp 27	1.02	0.
	1.61	9.1
Asn 27	1.02	28.5
	1.86	0.
Ser 27	1.02	29.5
	1.87	0.

Electron density difference maps are not presented here (ref. 27) but provide further evidence that some of the electron density changes resulting from proton transfer in the hydrogen bond (Tables 1 and 2) are of nonlocal character. Therefore, it appears to be necessary to describe the local electron densities in terms of electron distributions, well equilibrated on a larger area, possibly on the whole active site. We have designated such electron distributions as "hydrogen bond adapted".

TABLE 3

Calculated free energy changes for the inhibitor binding. Units: kcal/mole.

		ΔG	$\Delta \Delta G$	$\Delta \Delta G_{Exp.}$
E WT	-> E Asn	-7.9		
E WT : MTX	-> E Asn : MTX	-10.0	2.1	1.8
E WT	-> E Ser	-7.1		
E WT : MTX	-> E Ser : MTX	-11.9	4.8	4.4

Finally, in Tables 3 and 4 the resulting free energy changes for bound MTX and H₂F, respectively, are given. It is encouraging that differences in the charge stabilization by the enzyme active site can be described consistently within the ab initio formalism. Correct representation of this interaction then leads to better description of free energy changes by statistical mechanics.

TABLE 4
Calculated free energy changes for the substrate binding. Units: kcal/mole

			ΔG	$\Delta\Delta G$	$\Delta\Delta G_{Exp.}$
E WT	— >	E Asn	-7.9		
E WT :	— >	E Asn : H ₂ F	-10.3	2.4	2.2
E WT	— >	E Ser	-7.1		
E WT : H ₂ F	— >	E Ser : H ₂ F	-10.4	3.3	2.8

CONCLUSIONS

We have applied a novel methodology to determine binding constants of natural substrate dihydrofolate and methotrexate inhibitor to the three E. Coli DHFR enzymes: wild type, Asn27 and Ser27 mutants.

Use of ab initio calculated, "hydrogen bond adapted" charge parameters for Coulomb potential in the thermodynamic integration method implemented with molecular dynamics gives encouragingly good agreement with experiment. Our parametrization of the Hamiltonian seems to lead to linear time dependency of the free energy changes. Also, by taking into account all protein active site amino acid residues participating in the hydrogen bonding network with the ligand molecule on the ab initio level we were able to describe the subtle energetic differences between the hydrogen bonded mutant enzyme complexes with ligands and salt linked wild-type DHFR complex. This approach seems to be promising for general use in description of free energy changes upon binding in protein active sites where ligands often form hydrogen bonded networks with the protein.

ACKNOWLEDGEMENTS:

Our thanks are due to E.J. Villafranca (La Jolla) for providing us with the X-Ray coordinates. We appreciate the useful discussions with W.F. van Gunsteren (Groningen) and K. Mueller (Basel), and D. Turk (Martinsried) for help with molecular graphics. We would also like to thank KOMPAS Yugoslavia Travel Agency Computer Center (Mr. Pepelnjak - Manager) for generous allotment of computer time on their Vax 8650 system.

REFERENCES

1. J.E. Gready, Adv. in Pharmacol. and Chemother., 17 (1980) 37-102.
2. J.M. Blaney, C. Hansch, C. Silipo and A. Vittoria, Chem. Rev., 84 (1984) 333-407.
3. B.R. Baker, Cancer Chemother.Rep., 4 (1959) 1-10.

4. K. Hood and G.C.K. Roberts, *Biochem. J.*, **171** (1978) 357-366.
5. K. Taira and S.J. Benkovic, *J. Med. Chem.*, **31** (1988) 129 - 137.
6. H.B.Schlegel, M.Poe and K. Hoogsteen, *Mol. Pharmacol.*, **20** (1981) 154-158.
7. U.C.Singh, *Proc.Natl.Acad.Sci. USA*, **85** (1988) 4280-4284.
8. U.C.Singh and S.J.Benkovic, *Proc.Natl.Acad.Sci.USA*, **85** (1988) 9519-9523.
9. J.T. Bolin, D.J. Filman, D.A. Mathews, A.C. Hamlin and J. Kraut, *J. Biol. Chem.*, **257** (1982) 13650-13662.
10. D.K. Stammers, J.N. Champness, C.R. Bedell, J.G. Davis, E. Eliopoulos, A.J. Geddes, D. Ogg and A.C.T. North, *FEBS Lett.*, **218** (1987) 178-184.
11. C. Oeffner, A. D Arcy and F.K.Winkler, *Eur. J. Biochem.*, **174** (1986) 1123-1128.
12. E.E.Howell, J.E.Villafranca, M.S.Warren, S.J. Oatley and J.Kraut, *Science*, **231** (1986) 1123-1128.
13. R.E.London, E.E.Howell, M.S.Warren, J.Kraut and R.L. Blakely, *Biochemistry*, **25** (1986) 7235-7243.
14. A. Warshel, *Acc. Chem. Res.*, **14** (1981) 284.
15. A.J.Russell and A.R.Fersht, *Nature*, **328** (1987) 496-500.
16. A. Warshel and S.T.Russell, *Quart. Rev. Biophys.*, **17** (1984) 283-422.
17. T. Solmajer, *Stud. Biophys.*, **93** (1983) 149-159.
18. M. Hodoscek and T. Solmajer, *Theor. Chim. Acta*, **66** (1985) 395-403.
19. T.Solmajer, M.Hodoscek and D. Hadzi, *Int.J. Quant.Chem.*, **23** (1983) 945-952.
20. J.P.M.Postma, H.J.C.Berendsen and J.R. Haak, *Faraday Symp. Chem.Soc.*, **17** (1982) 55.
21. P. A. Bash, U.C. Singh, R. Langredge and P.A.Kollman, *Science*, **236** (1987) 564-568.
22. P.A.Bash, U.C.Singh, F.K.Brown, R.Langredge and P.A.Kollman, *Science*, **235** (1987) 574-576.
23. R. Poirier and M. Peterson, *Monstergauss*, Dept.of Chem. University of Toronto, Toronto, Ontario, Canada. 1985.
24. W.F. van Gunsteren, *Gromos Reference Manual*, Biomos B.V. Nijenborgh 16 Groningen, The Netherlands, 1987.
25. H.J.C. Berendsen, J.P.M. Postma, W.F.van Gunsteren and J.Hermans, *Intermolecular Forces*, (B. Pullman, Ed.), Reidel, Dordrecht, 1981, p. 331.
26. T.P. Straatsma and H.J.C. Berendsen *J. Chem. Phys.*, **89** (1988) 5876-5886.
27. T. Solmajer, submitted for publication.
28. A. Cross, *Chem Phys. Lett.*, **128** (1986) 198-202.
29. M. Hodoscek, D. Kocjan and T. Solmajer, *Croat. Chim. Acta*, **57** (1984) 65-74.
30. M. Hodoscek, D. Hadzi and T. Solmajer, *J. Mol. Struct. (Theochem.)*, **183** (1989) 371-379.
31. J.T. Chen, R.I. Mayer, C. A. Fierke and S.J. Benkovic, *cell. Biochem.*, **29** (1985) 73 - 82.

MODELLING BASE SEQUENCE EFFECTS AND TRANSITIONS IN DNA

R. LAVERY¹, M. PONCIN¹ and H. SKLENAR²

¹Institut de Biologie Physico-Chimique, 13 rue Pierre et Marie Curie, Paris 75005, France

²Central Institute of Molecular Biology, Academy of Sciences of the GDR, Robert Rössle Str. 10, Berlin Buch, DDR-1115

SUMMARY

The use of internal and helicoidal coordinates for modelling biopolymers is illustrated for the case of double stranded deoxyribonucleic acid. This approach enables conformational changes, junctions and fine base sequence effects to be studied in a controlled way. A new approach to dealing with solvent effects is also discussed.

INTRODUCTION

The study of the properties and interactions of deoxyribonucleic acid is a subject of continually growing importance. Over the last 10 years, developments in oligonucleotide synthesis and subsequent single crystal x-ray studies have shown DNA structure to be strikingly more complex than was previously suspected. Fine structure as a function of base sequence is now known to play an important role in protein-nucleic acid and drug-nucleic acid recognition and, in particular, long range effects, mediated by sequence dependent axis curvature, have come to light. Moreover, the biological role of unusual conformations such as the left-handed Z form, cruciforms and triple stranded segments has been demonstrated.

Reliable computer modelling of large fragments of DNA is thus of considerable importance. However, such studies present a number of problems which have not yet been fully solved. Firstly, in common with studies of other biopolymers, DNA modelling requires the treatment of systems containing 1000 atoms or more. Secondly, the poly-ionic nature of DNA makes it remarkably sensitive to environmental factors. Finally, many interesting biological processes concerning DNA clearly require its flexibility to be taken into account and often involve large conformation changes.

Our approach to these problems has been, firstly, to develop a series of algorithms (refs. 1-4) which enable DNA fragments to be modelled with less variables than are required by the classical molecular mechanics or dynamics procedures. Through the use of helicoidal and internal variables we are now able to study a DNA duplex with 10 times fewer variables or, in the case of imposed helical symmetry, 100 times fewer variables than would otherwise be necessary. Moreover, the direct use of helicoidal variables enables us to provoke conformational changes in a fully controlled way and thus to investigate their energetics.

The present article summarizes these methodological developments, concentrating on our most recent algorithm (ref. 4) termed JUMNA (Junction Minimisation of Nucleic Acids), as well as describing some of the applications that we have been able to carry out for duplex DNA. We also describe the first step in the next stage of our study which is aimed at achieving a more realistic model of the solvent and counterion environment which surrounds DNA within the cell. This development, termed FIESTA (Field Integrated Electrostatic Approach), enables the crude dielectric damping usually employed in simulation studies to be replaced by a greatly improved representation of the influence of a continuum dielectric surrounding the van der Waals envelope of a macromolecule.

METHODS

Energy formulation

The first step in modelling DNA behavior is to formulate an energy functional which will describe as well as possible its internal stability. Computational limits are nevertheless imposed on this functional by the size of the systems we will treat which often contain more than 1000 atoms. Within these limits we have tried to develop a functional which is as precise as possible. We have paid particular attention to electrostatic interactions and also to hydrogen bond formation which plays an important role both within DNA and for its interaction with other molecules. The functional we employ, termed "Flex" (refs. 2,3,5) is given below:

$$\begin{aligned}
 E = & \sum Q_i Q_j / \epsilon(R) R_{ij} + \sum (-A_{ij} / R_{ij}^6 + B_{ij} / R_{ij}^{12}) \\
 + & \sum [\cos \theta (-A_{ij}^{HB} / R_{ij}^6 + B_{ij}^{HB} / R_{ij}^{12}) \\
 + & (1 - \cos \theta) (-A_{ij} / R_{ij}^6 + B_{ij} / R_{ij}^{12})] \\
 + & \sum V_s / 2 (1 \pm \cos N_s^T s) + \sum V_a (\omega_a - \omega_a^0)^2
 \end{aligned}$$

This expression consists of a series of pairwise additive terms. The first term represents the electrostatic energy, calculated as a sum of interactions between atomic monopoles Q_i and damped by a dielectric function $\epsilon(R)$ dependent on the interaction distance. The following three terms represent dispersion-repulsion interactions calculated with a 6-12 dependence using, as a basis, the parameter set developed by Zhurkin et al (ref. 6). Hydrogen bonds are dealt with by the latter two of these terms, which take into account their angular dependence. These parameters were determined by ab initio SCF calculations on model systems for each type of bond. All of these terms are summed over pairs of atoms separated by at least three chemical bonds. The last two terms represent the distortion energy associated with torsion angles τ_s (including anomeric effects) and valence angles ω_a which were derived from experimental values or again obtained by ab initio calculations on model systems.

The atomic monopoles we employ are calculated by the Huckel-Del Re technique with a set of parameters specially developed for the nucleic acids. These parameters were derived by fitting the monopole potentials and fields on surface envelopes surrounding the nucleic acid sub-units to the exact values obtained by ab initio SCF calculations (ref. 5). Particular care was taken to reproduce correct charge distribution between these sub-units by explicit ab initio calculations on nucleosides and S-P-S and P-S-P fragments. The dielectric function we have used in the studies presented here is based on the form proposed by Hingerty et al (ref. 7). We have reformulated this function as shown below so that it is possible to vary both the plateau value of the dielectric reached at long distance (D) and the slope of the sigmoidal segment of the function (S).

$$\epsilon(R) = D - (D-1)/2 [(RS)^2 + 2RS + 2] \exp(-RS)$$

Trial calculations on the different conformational forms of DNA showed the original distance dependence proposed by Hingerty, (corresponding to a plateau value of 80 and a slope of 0.356) with net phosphate charges reduced to $-0.5e$ led to good results for high damping conditions, notably for the Z conformation. However, for the A and B conformations a lower slope of 0.16 led to the best overall results.

Using internal coordinates

The simplest way to define the conformation of a molecular system for computer study is in terms of cartesian coordinates, leading to a total of $3N$ variables for N atoms. Although this technique is used in conventional molecular mechanics it leads to considerable minimisation problems for biological macromolecules which will generally be represented by several thousand variables. If one wants to reduce the time necessary for optimisation and also to reduce the problems caused by local minima in very high dimensional spaces the only possibility is to treat fewer variables and to find variables which describe molecular motion more effectively.

This can be done in a chemically meaningful way by adopting internal coordinates: bond lengths, valence angles and dihedral angles. Directly using such variables simplifies the treatment of the conjugated motion of atoms, corresponding, for instance, to rotation around a single bond. A further advantage of this approach is the possibility of fixing any of the internal variables during minimisation, notably bond length variations. This can be a powerful way of simplifying the model since such variations are generally associated with considerably higher force constants than either valence or torsion angles.

By dramatically reducing the number of variables necessary to represent a macromolecular system, this approach can be used to study much larger conformational changes than would otherwise be possible. This formulation can also be extended to the description of polymers and, in particular DNA, directly in terms of helicoidal parameters. This gives us the possibility of easily controlling important aspects of their conformation and, as we shall show, of either imposing helicoidal symmetry or of inducing chosen types of conformational distortion.

However, one problem arises when using internal coordinates in the presence of geometrical constraints implied by ring closure or helicoidal parameters. In these cases, internal coordinates represent an over-description of the molecular system. Certain variables become linked by equations which are usually coupled, non-linear and difficult to solve analytically. This problem can nevertheless be overcome and we will illustrate this in the case of our most recent algorithm termed JUMNA.

JUMNA (Junction Minimisation of Nucleic Acids)

The aim of our most recent methodology, JUMNA (ref. 4), is to combine the control over the helicoidal parameters of DNA with an effective treatment of the constraint problem. In order to achieve this we divide the nucleic acid fragment studied into 3'-monophosphate nucleotides which are positioned with respect to an axis using helicoidal parameters. These parameters, combined with backbone torsions and sugar valence angles become the variables of our model. A constraint minimizer is then used to satisfy a total of 4 constraints per nucleotide: the sugar ring closure distance, a backbone closure distance between O5' and C5', and two valence angles P-O5'-C5' and O5'-C5'-C4'.

This approach has several notable advantages. Firstly, the "junctions" between consecutive nucleotides do not have to be closed for the starting structure and, consequently, it is not necessary to find the appropriate nucleotide conformations corresponding to a chosen set of helicoidal parameters before beginning energy optimisation. This greatly simplifies the investigation of unusual and irregular nucleic acid structures for which no conformational data exists. Secondly, "junctions" between nucleotides can open during optimisation allowing for passage between conformational states which would otherwise be separated by large energy barriers. Thirdly, it is not necessary to develop and solve the complicated equations describing the "closure" of the constrained system. Fourthly, the direct use of helicoidal variables allows overall deformations such as stretching, twisting and bending to be carried out in a controlled way. It can finally be remarked that it is easy to impose symmetry on the DNA fragments studied by simply using a single variable in place of the corresponding helicoidal parameters of a set of nucleotides. In this way, our present version of JUMNA allows the imposition of mononucleotide or dinucleotide repeat symmetry and can treat up to 4-stranded systems.

The JUMNA algorithm thus allows easy construction and energy optimisation of both regular and irregular DNA oligomers. The small number of variables necessary, the speed of optimisation and the control over the final conformation make this approach very interesting for modelling a wide variety of biologically important problems involving the nucleic acids. It is finally remarked that we have also recently developed an algorithm

(CURVES, refs. 8,9) which allows the optimal helical axis and a full helicoidal parameter set to be obtained for irregular DNA conformations. This tool is of considerable help in interpreting sequence effects within irregular structures in a clear and rigorous way.

RESULTS AND DISCUSSION

(i) Deformation of Symmetric DNA duplexes

We begin by looking at the flexibility of energy optimised B-DNA and A-DNA duplexes with respect to various types of regular distortion. We will consider three changes with respect to helicoidal parameters: twist or winding angle (WDG), base pair rise (ZSH) and base pair displacement along the dyadic axis (XSH), and also one backbone variable, sugar pucker phase angle (PHA). The first two of these distortions are represented graphically in figures 1 and 2 for the B and A conformations of DNA. In each case two homopolymeric sequences have been considered, poly(dG).poly(dC) and poly(dA).poly(dT). From these results it is immediately obvious that the B conformation is much more flexible and this fact is confirmed for the other types of distortion in table 1 where the sharpness of the distortion energy curves is measured by their half-width at 0.5RT (0.3 kcal/mole) above the minimum. It can also be seen that, for B-DNA, the AT sequence is also somewhat stiffer than the GC sequence while for A-DNA sequence effects are less marked.

TABLE 1

Half-width of distortion curves 0.5RT above the minimum

Form	Sequence	WDG(°)	ZSH(Å)	XSH(Å)	PHA(°)
B	GC	4.2	0.4	1.5	14.0
B	AT	3.3	0.2	1.1	7.0
A	GC	0.3	0.05	0.11	2.0
A	AT	0.4	0.07	0.09	2.0

In geometrical terms, the deformation of the two allomorphic forms studied occurs in different ways. In B-DNA, helicoidal deformation is generally absorbed by coupled changes in other helicoidal parameters, while for A-DNA the adaption mainly

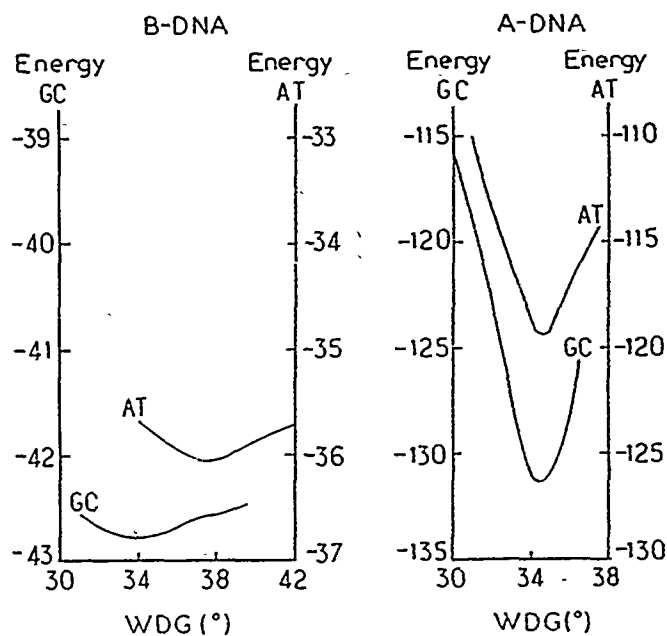


Figure 1. Deformation energy (kcal/mole) of B- and A-DNA with respect to Twist (WDG)

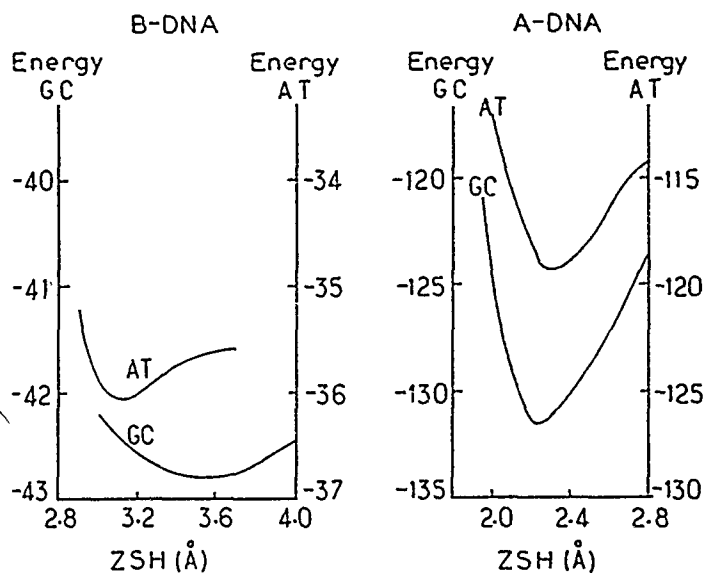


Figure 2. Deformation energy (kcal/mole) of B- and A-DNA with respect to rise (ZSH)

takes place through rearrangement of the sugar phosphate backbone. One exception to this occurs for base pair rise which, for A-DNA, leads to a rapid decrease in tilt as rise increases. Overall, it appears that the stiffness and the poor coupling of helicoidal parameter changes in the A conformation not only lead to its overall rigidity, but also diminish the influence of base sequence on its properties.

(ii) Variable correlations and the B-A transition

When the conformations resulting from the distortions discussed above are grouped together it is possible to detect a number of general correlations between different variables. In the case of B-DNA we find 6 linear correlations: $(C3'-C4') : WDG$, $(C3'-C4') : -(O3'-P)$, $(C3'-C4') : GLY$ (glycosidic angle), $(O3'-P) : -XSH$, $ZSH : TLT$ (tilt angle) and $XSH : -WDG$. Of these, the first three have been observed crystallographically for B-DNA oligomers (refs. 10,11). In the case of A-DNA we find 4 correlations: $(C3'-C4') : GLY$, $(P-O5') : -(C5'-C4')$, $ZSH : -TLT$ and $TLT : -GLY$. Again, the first of these correlations has been observed crystallographically. Considering the variety of ways in which the deformations we have imposed could be absorbed it is encouraging to find these agreements with experimental data. Overall, this would seem to indicate that our reduced variable model of DNA coupled with the "Flex" energy formulation is relatively successful in describing DNA mechanics.

It is clear from the results presented that the A and B conformations truly represent two distinct conformational forms of the DNA double helix. This is particularly clear for the correlations discussed above which are generally quite different for each allomorphic form. The one exception to this rule occurs for the correlation between sugar puckering (closely linked to the $C3'-C4'$ dihedral angle) and the glycosidic torsion. These variables are coupled in the same way for both the A and B forms, despite the fact that the optimum value of each variable is very different in the two allomorphs. This observation led us to search for a simple transition pathway from the B to the A form. While forced sugar repuckering did not lead to an appropriate conformational change, reduction of the glycosidic angle resulted in an easy transition from the B to the A form with virtually no energy barrier, maintaining a helically symmetric conformation at each step. This pathway, illustrated in figure 3, is one example

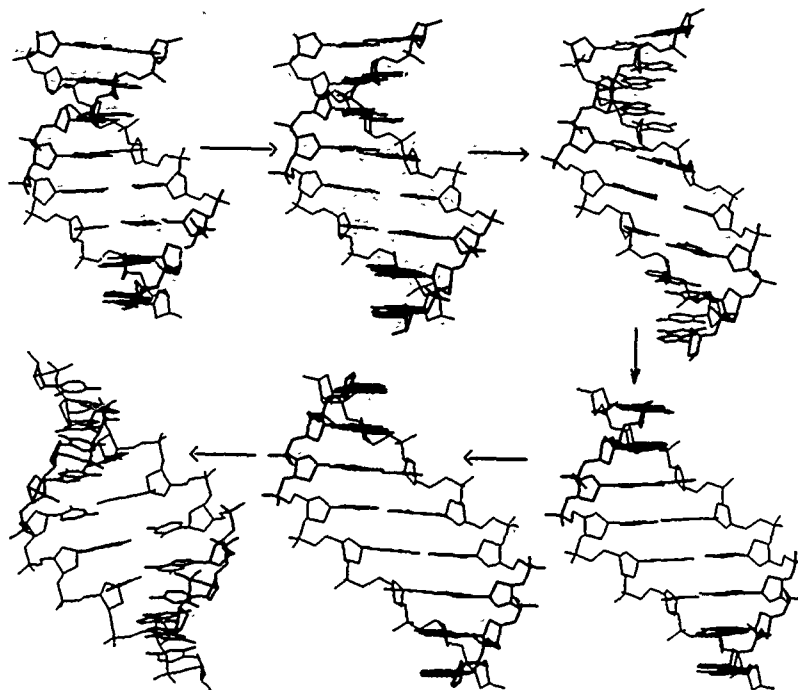


Figure 3. B (top left) to A (bottom left) transition pathway by glycosidic angle control

of how control over the helicoidal and internal variables of DNA enables us to investigate very large conformational changes in a controlled way and with only limited computational expense.

(iii) B-Z junctions and a possible transition mechanism

Since experimental evidence suggests that the B to Z transition takes place via a junction moving along the double helix, we began our studies in this area by creating a B-Z junction model (ref. 4). Starting from two 3-nucleotide pair fragments in relaxed B and Z conformations, we were able to profit from the helicoidal variable construction used by JUMNA to form two types of junction without any base pair disruption. The two junctions, corresponding to 5'(CpG)3' and 5'(GpC)3' nucleotides at the B-Z interface (the overall sequence being respectively CGCGCG and GCGCGC), both required a 6Å shift of the Z fragment towards the minor groove side of the B duplex. After energy optimisation

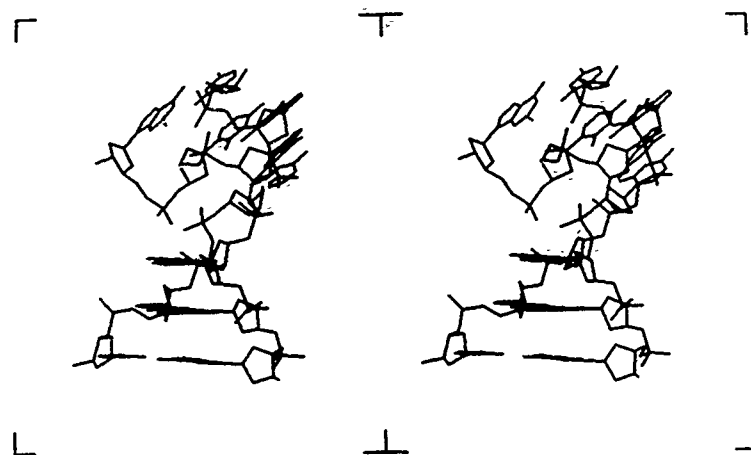


Figure 4. CpG B-Z junction showing kink towards major groove of the B segment

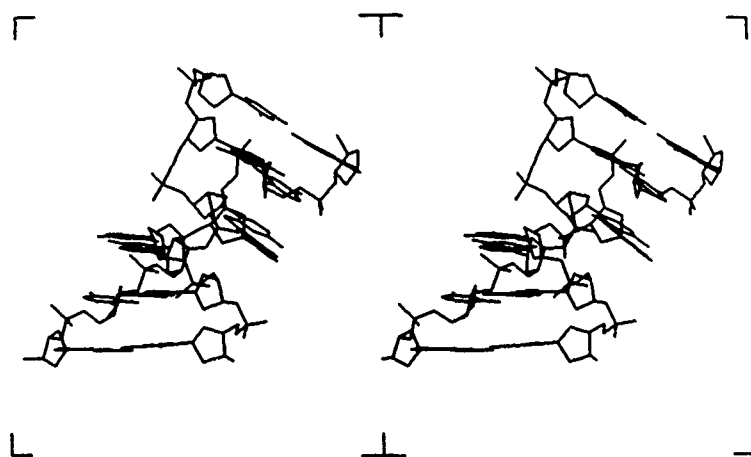


Figure 5. GpC B-Z junction showing kink towards minor groove of the B segment

these junctions led to very different conformations. While the CpG junction resulted in a large rise at the B-Z interface and a kink of roughly 40° towards the major groove side of the B duplex, the GpC junction kinked by almost 30° in the opposite direction and was able to insert the 5'-3' backbone of the Z segment into the minor groove of the B helix. As we have discussed in an earlier publication (ref. 4) the reason for the junction kinks can be explained by the rotation of the base pairs around their long axis (right-handed with respect to the G \rightarrow C vector) during the B-Z transition. This rotation, coupled with fact that the cytidine nucleotide remains anti, leads to a stretched sugar-phosphate linkage at the interface which can be relaxed by kinking in the directions observed. Comparing the energy of these structures with the mean energy of a 6-nucleotide pair B or Z helix gave the formation energies as 19.3 kcal/mole for CpG and only 10.3 kcal/mole for GpC. These structures are represented in the stereodiagrams figures 4 and 5.

Starting from these structures it was possible to construct a transition pathway which led the uppermost B-DNA pair to fully rotate and become part of the Z fragment. This conformational change, which was accomplished by profiting from the possibility of locking the propeller twists within the JUMNA algorithm, is illustrated in figure 6. It should be noted that this pathway did not involve any base pair opening and corresponded to an energy barrier of roughly 25 kcal/mole. The fact the junctions at either end of this transition pathway were kinked in opposite directions turned out to favour the transition mechanism, since stacking energy between the rotating pair at the interface and the Z segment could be regained more rapidly as the kink direction changed than if the junction had remained straight.

(iv) Sequence effects on DNA conformation

The studies we have carried out to date (refs. 12,13) have led to the detection of important and reproducible base sequence effects on the conformations of B-DNA, A-RNA, Z-DNA and Z-RNA. For a detailed discussion of the effects within the Z conformation and the observed correlations with chemical reactivity studies, the reader is referred to another article in the present volume (ref. 14). We will limit ourselves here to the latest results involving the B conformation of DNA.

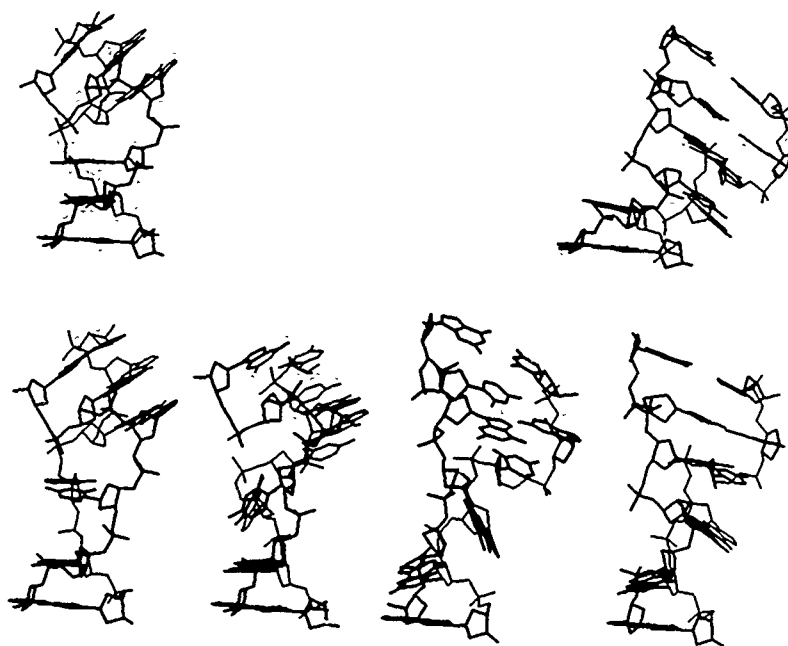


Figure 6. B-Z transtion pathway involving a passage from a CpG junction (top left) to a GpC junction (top right) by the rotation of the B pair at the interface.

Our initial studies of B-DNA led to a number of striking sequence effects involving changes in twist of almost 12° and changes in rise of more than 1\AA . Tilt also varied widely with base composition notably yielding a large negative value for homopolymeric G tracts, while propeller twist was large and positive for A tracts. The twist results also appeared to show a good correlation with the values deduced by Kabsch, Sander and Trifonov (ref. 15) by fitting to experimental data. However, following deformation studies of these optimal geometries it was found that other minima, which we will term B^* , corresponding to a partial transition to the A conformation existed and were, with one exception, more stable than the true B forms.

TABLE 2

Conformation and energy changes (kcal/mole) upon passing from the B to the B* conformation as a function of base sequence

Sequence	Sugar B	Twist B	Sugar B*	Twist B*	$\Delta E(B \rightarrow B^*)$
(GC)n	C2'endo	GpC>CpG	G:C2'endo C:O1'endo	GpC<CpG	1.3
(AT)n	C2'endo	ApT>TpA	A:C2'endo T:O1'endo	ApT<TpA	0.7
(GT)n	C2'endo	GpT>TpG	C:O1'endo Others:C2'endo	GpT<TpG	3.8
(CT)n	C2'endo	CpT=TpC	C:O1'endo Others:C2'endo	CpT<TpC	1.8
(AA)n	C2'endo		A:C2'endo T:C3'endo		3.0
(GG)n	C2'endo		G:C2'endo C:C3'endo		-1.6

The results listed in table 2 show that these new conformations resulted from a transition of most of the pyrimidine sugars to north puckers (O1'endo or C3'endo) and also involved inversions in the dinucleotide features we had previously observed. This is notably the case for twist variations within dinucleotide repeat sequences as illustrated in the table. It was also discovered that these new intermediate conformations were stiffer with respect to helicoidal deformation, in line with the results found for A-DNA and discussed in section (i). We believe that the existence of these forms is linked to the fact that the A conformation is intrinsically more stable than the B conformation using our simplified dielectric model. These results made it clear, at least for the highly flexible B form, that it was necessary to improve the modelling of the environment surrounding DNA and our first step in this direction is described in the following section.

(v) FIESTA: a new approach to modelling solvent effects

The basis of the FIESTA (Field Integrated Electrostatic Approach) is contained in the classical electrostatic expression which describes the work necessary to assemble a charge distribution within a dielectric medium in terms of the field (\vec{E}) and the dielectric displacement (\vec{D}) generated.

$$dw = \bar{E} \bar{D} / 8 \pi dt = \epsilon \bar{E}^2 / 8 \pi dt$$

As shown, under normal conditions where neither dielectric saturation or high frequency fields are involved, \bar{D} may be written as $\epsilon \bar{E}$. This equation can be used to calculate the electrostatic energy of a system of charges as the following examples show.

EXAMPLE 1: The interaction of two point charges q_1 and q_2 placed at points r_1 and r_2 respectively within a continuum dielectric.

The field at an arbitrary point r can be written as,

$$\bar{E}(r) = q_1 (\bar{r} - \bar{r}_1) / \epsilon (r - r_1)^3 + q_2 (\bar{r} - \bar{r}_2) / \epsilon (r - r_2)^3$$

The work necessary to assemble this system is thus,

$$\begin{aligned} w(r_1, r_2) &= (1/2 \epsilon) (q_1^2 + q_2^2) \int dr / r^2 \\ &\quad + (q_1 q_2 / \epsilon) \int \{ (\bar{r} - \bar{r}_1) (\bar{r} - \bar{r}_2) / [(r - r_1)^3 (r - r_2)^3] \} dt \\ &= \text{divergent self-energy} + q_1 q_2 / \epsilon r_{12} \end{aligned}$$

which as we can see is simply the usual Coulomb expression for the energy plus a divergent term associated with each individual charge that can be ignored.

EXAMPLE 2: The transfer of an ion of radius 'a' from vacuum to a dielectric medium.

In this case, we can write the work necessary for the transfer as,

$$w = w - w_0 = 1/8 \pi \int dw (\bar{E} \bar{D} - \bar{E}_0 \bar{D}_0) r^2 dr$$

where \bar{E} and \bar{D} refer to the situation within the dielectric and \bar{E}_0 and \bar{D}_0 are the vacuum values. Since the dielectric constant is unity in vacuum $\bar{D}_0 = \bar{E}_0$. Also $\bar{E} = \bar{E}_0 / \epsilon$, thus $\bar{D} = \bar{E}_0$ and,

$$\begin{aligned} w &= q^2 / 2 (1/\epsilon - 1) \int \bar{E}_0^2 r^2 dr \\ &= -1/2 (1 - 1/\epsilon) (q^2 / a) \end{aligned}$$

which is the well known Born formula for ionic solvation.

The same approach, after a number of reasonable approximations, can be extended to calculate the energy of a molecule within a continuum dielectric. We will firstly represent the molecule as a distribution of n atomic point charges q_i surrounded by an envelope formed from the superposition of van der Waals atomic spheres (radii a_i). The total volume inside the envelope is divided up into atomic volumes C_i by taking into account the overlap of the van der Waals spheres of chemically bound atoms. These volumes are termed CPK atomic volumes since

they correspond to the way a CPK molecular model is built up. It will presently be assumed that the dielectric within the molecular envelope is unity. If we now assume that the dielectric displacement (or field) generated by the charge distribution is not affected by the presence of the dielectric boundary then we can write the electrostatic potential energy of this system as a sum of three terms:

$$V(r_1, r_2, \dots, r_n) = V_0 + V_1(r_1, r_2, \dots, r_n) + V_2(r_1, r_2, \dots, r_n)$$

The first of these terms is a zero point energy, which is simply the sum of the Born transfer energies for each individual atom,

$$V_0 = -1/2 (1-1/\epsilon) q_i^2/a_i$$

The second term corresponds to the Coulomb interaction energies of the point charges within the dielectric medium,

$$V_1 = q_i q_j / (\epsilon r_{ij})$$

and the third term corresponds to the field energy associated with the vacuum envelope surrounding the charge distribution,

$$V_2 = f v / 8 \pi (1-1/\epsilon) \int \bar{E}^2 dv$$

Note that this term contains the only adjustable parameter of the method, $f v$ a volume factor, which is used to correct for the fact that discrete solvent molecules cannot perfectly fill the space outside the van der Waals molecular envelope.

In order to evaluate the field integral we will separate the field contributions at any given atom i into the field \bar{I}_i created by the atom itself, the field \bar{J}_i created by atoms b chemically bound to i , and the field \bar{K}_i due to the remaining atoms.

$$\bar{E}_i = \bar{I}_i + \bar{J}_i + \bar{K}_i$$

After squaring and regrouping terms we can see that,

$$\int \bar{E}_i^2 = \int \bar{I}_i^2 + \int (2\bar{I}_i + \bar{J}_i) \cdot \bar{J}_i + \int (2\bar{I}_i + 2\bar{J}_i + \bar{K}_i) \cdot \bar{K}_i$$

The first of these integrals is a divergent self-energy which may be ignored as in the example shown above. The second integral, which we will term A_i must be evaluated. The final integral can be simplified if we assume that the field from the more distant atoms \bar{K}_i is homogeneous within the CPK volume of atom i . In this case we can calculate a simple volume average,

$$\bar{F}_i = 2 \int (\bar{I}_i + \bar{J}_i) dv / C_i$$

and rewrite the last integral as,

$$C_i (\bar{F}_i + \bar{K}_i) \cdot \bar{K}_i$$

The volume integrals A_i and \bar{F}_i unfortunately cannot be solved analytically, however it is possible to separate the atomic

charges they contain to generate a set of simpler integrals,

$$A_i = \sum_b q_b [2q_i S_{ib} + \sum_{b'} q_{b'} T_{bb'}]$$

$$\bar{F}_i = 2 (q_i \bar{Z}_i + \sum_b q_b \bar{U}_{ib})$$

The volume integrals S, T, \bar{Z} and \bar{U} only depend on the geometrical positioning of the neighbours b chemically bound to atom i .

$$S_{ib} = \int (\bar{r} - \bar{r}_i) \cdot (\bar{r} - \bar{r}_b) / [(r - r_i)^3 (r - r_b)^3] dv$$

$$T_{bb'} = \int (\bar{r} - \bar{r}_b) \cdot (\bar{r} - \bar{r}_{b'}) / [(r - r_b)^3 (r - r_{b'})^3] dv$$

$$\bar{Z}_i = \int [(\bar{r} - \bar{r}_i) / (r - r_i)^3] dv / C_i$$

$$\bar{U}_{ib} = \int [(\bar{r} - \bar{r}_b) / (r - r_b)^3] dv / C_i$$

These integrals can be calculated numerically and, with the use of standard bond lengths and angles, may be built up into an integral data base for each type of atomic center that will be required (defined by the type of central atom, the types of its bound neighbours and their spacial positions). Using these standardized integrals results in a error of roughly 5% compared to using the exact molecular geometry. Several other refinements can be made to this methodology, notably a slight modification of the interaction distance r_{ij} used to evaluate the fields \bar{K}_i of distant atoms which improves the quality of the volume average introduced in the \bar{F}_i integral and also a correction to allow for overlap between unbound atoms, notably those forming hydrogen bonds. It is also remarked that the Born energies of the V_0 term may be made over the full atomic spheres rather than the CPK volumes C_i by a modification of the volume taken into account in the $T_{bb'}$ integrals described above.

The major simplification introduced in the FIESTA model involving the removal of the dielectric interface effects on the dielectric displacement clearly needs to be justified by tests on realistic systems. The first such check was carried out for the case of a number of charges contained within a large spherical cavity (using a close packed array of small spheres to mimic a molecular envelope). This calculation led to an excellent agreement with the exact solution obtained by Kirkwood (ref. 16). In the case of real molecules it is possible to consider a refined parameterisation of the atomic van der Waals radii, however as the results in table 3 show, simply using Pauling radii and a volume factor $fv=1.25$ (with our usual Huckel-Del Re charges from the "Flex" parameterisation) it is already possible to obtain an

excellent agreement with the experimental hydration enthalpies for the nucleic acid bases. In addition the FIESTA value for dimethyl phosphate hydration (in the gt conformation), -92.8 kcal/mole, falls close to the value of -85.4 kcal/mole (ref. 17) obtained by a more refined discrete water + continuum calculation (which however used a rather different charge distribution obtained from CNDO studies).

TABLE 3

Comparison of FIESTA and experimental hydration enthalpies (refs. 18,19) for nucleic acid subunits (kcal/mole)

Molecule	FIESTA	Experimental
Guanine	-34.1	-35.4 \pm 2.6
Adenine	-24.7	-22.6 \pm 0.6
Cytosine	-29.2	-29.5 \pm 1.0
Thymine	-27.3	-23.7 \pm 0.7

Overall, it would appear that the FIESTA model can be used to greatly improve the treatment of solvent effects within macromolecular simulations. It can be shown to correctly reproduce many qualitative features of the dielectric environment (short range repulsion between charges of opposite sign, correct asymptotic behavior of the apparent dielectric for charges buried within a molecule or close to the solvent interface, forces on buried charges leading them towards the solvent, etc) which are absent in the simple dielectric models generally employed today. From a practical point of view it is important to note that the FIESTA energy and its analytic derivatives can be evaluated rapidly, even for macromolecular systems, and thus can easily be incorporated in a general conformational energy optimisation procedure.

CONCLUSIONS

This article summarizes the steps that we have been able to take towards a more refined modelling of the properties of the nucleic acids. The combined use of internal and helicoidal coordinates within the JUMNA algorithm form an excellent basis for investigating both fine base sequence effects, flexibility and

large conformational changes and have already allowed us to study a number of biologically interesting problems. Coupling this approach with the new solvent treatment, FIESTA, that we have described should now make it possible to take another step towards a truly realistic simulation of macromolecular behaviour.

REFERENCES

- 1 H.Sklenar, R.Lavery and B.Pullman, The flexibility of the nucleic acids: (I) "SIR", a novel approach to the variation of polymer geometry in constrained systems, *J. Biomol. Struct. Dynam.* 3 (1986) 967-987.
- 2 R.Lavery, H.Sklenar, K.Zakrzewska and B.Pullman, The flexibility of the nucleic acids: (II) The calculation of internal energy and applications to mononucleotide repeat DNA, *J. Biomol. Struct. Dynam.* 3 (1986) 989-1014.
- 3 R.Lavery, I.Parker and J.Kendrick, A general approach to the optimization of the conformation of ring molecules with an application to valinomycin, *J. Biomol. Struct. Dynam.* 4 (1986) 443-461.
- 4 R.Lavery, Junctions and bends in nucleic acids: A new theoretical modelling approach, in W.K.Olson, R.H.Sarma, M.H.Sarma and M.Sundaralingam (eds.) *Structure and Expression Vol.3 DNA Bending and Curvature* Adenine Press, New York 1988 pp.191-211.
- 5 R.Lavery, K.Zakrzewska and A.Pullman, Optimized monopole expansions for the representation of the electrostatic properties of the nucleic acids, *J.Comp.Chem.* 5 (1984) 363-373.
- 6 V.B.Zhurkin, V.I.Poltev and V.L.Florent'ev, Atom-atom potential functions for conformational calculations of nucleic acids, *Molekulyarnaya Biologiya* 14 (1980) 882-895.
- 7 B.E.Hingerty, R.H.Richie, T.L.Ferrell and J.E.Turner, Dielectric effects in biopolymers: The theory of ionic saturation revisited, *Biopolymers* 24 (1985) 427-439.
- 8 R.Lavery and H.Sklenar, The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids, *J. Biomol. Struct. Dynam.* 6 (1988) 63-91.
- 9 R.Lavery and H.Sklenar, Defining the structure of irregular nucleic acids: Conventions and principles, *J. Biomol. Struct. Dynam.* 6 (1989) 655-667.
- 10 R.E.Dickerson, M.L.Kopka and H.R.Drew, Structure correlations in B-DNA, in E.Clementi and R.H.Sarma (eds) *Structure and Dynamics: Nucleic acids and Proteins* Adenine Press New York (1982) pp 149-179.
- 11 R.E.Dickerson and H.R.Drew, Kinematic model for DNA, *Proc. Natl. Acad. Sci. (USA)* 78 (1981) 7318-7322.
- 12 B.Hartmann, B.Malfroy and R.Lavery, Theoretical prediction of base sequence effects in DNA. Experimental reactivity of Z-DNA and B-Z transition enthalpies, *J. Mol. Biol.* 207 (1989) 433-444.
- 13 B.Hartmann and R.Lavery, The conformation and stability of ribonucleic acids: Modelling base sequence effects in double stranded helices, *J. Biomol. Struct. Dynam.* (1989) in press.
- 14 B.Hartmann and R.Lavery, Theoretical prediction of base pair sequence effects in Z-DNA and Z-RNA, in the present volume.
- 15 W.Kabsch, C.Sander and E.N.Trifonov, The ten helical twist angles of B-DNA, *Nucl. Acids Res.* 10 (1982) 1097-1104.

- 16 J.G.Kirkwood, Theory of solutions of molecules containing widely separated charges with special application to zwitterions, *J. Chem. Phys.* 2 (1934) 351-361.
- 17 J.Langlet, P.Claverie, B.Pullman and D.Piazzola, Studies of solvent effects (IV) Study of hydration of the dimethyl phosphate anion (DMP^-) and of the solvent effects on its conformation, *Int. J. Quantum Chem. Quantum Biology Symp.* 6 (1979) 409-437.
- 18 I.K.Yanson, A.B.Teplitsky and I.F.Sukhodub, Experimental studies of interactions between nitrogen bases of nucleic acids, *Biopolymers* 18 (1979) 1149-1170.
- 19 S.Cabani and P.Gianni, Gas-liquid and solid-liquid phase equilibria in binary aqueous systems of nonelectrolytes, in H-J.Hinz (ed), *Thermodynamic Data for Biochemistry and Biotechnology*, Springer-Verlag Berlin (1986) pp 259-275.

DISCUSSION

PEPE - The electric field around ions is very strong. Did you take into account the effect of the electric field on the orientation of the water molecules ?

LAVERY - As you have seen the FIESTA model treats the solvent as a continuum. In consequence, no allowance for the effects of discrete waters close to the solute is made. If such effects turned out to be of prime importance then a continuum model would undoubtedly be a poor choice. It is thus perhaps encouraging to recall that, in the case of the dimethylphosphate anion, we have obtained a very good agreement with a more refined model which included an explicit first hydration shell. It would seem that, in this case, even the strong fields produced by a net charge of -1 can effectively be dealt with within the continuum treatment. One of our reasons for starting with this approach was after all that the Born model for ionic hydration works so well. We should also remark that FIESTA, despite its simplicity, certainly represents an important step in quality beyond simple distance dependent dielectric functions which are commonly used today.

KARPLUS - Why do you quote enthalpies of hydration for your model, I thought this treatment should lead to free energies ?

LAVERY - You are quite correct, in principle, the FIESTA model, like the Born ionic hydration model, leads to an electrostatic free energy. However, since the entropy enters this simple model through the temperature dependence of the dielectric constant we did not feel that this was likely to yield a good treatment of entropic factors. We therefore chose to assume, for the moment, that we are dealing with a model dielectric where this dependence is removed and fit the electrostatic term to enthalpies alone. This is also more coherent with the enthalpic contributions which we calculate for the other terms in our energy functional.

SOUMPASIS - There are many stereochemically personable propositions for the structure of B-Z junctions. But in the absence of any definitive experiments which may discriminate them, that they are a priori equiprobable is equally plausible. However, there is some evidence that single stranded regions are very probably

present (4-S bp's) on the basis of enzyme digestion studies. How does your junction model explain this fact ?

LAVERY - Our model not only suggests a stereochemically and energetically feasible conformation for B-Z junctions, but also leads to a possible transition pathway which is generally more difficult to find. As you suggest some studies indicate strand opening and enhanced chemical reactivity at the junctions has also been found. It is nevertheless difficult to distinguish between a true open state and a simple fragility at the junction region. Within our model, the lack of stacking between the base pairs at the interface will certainly weaken the junction zone and allow easier chemical attack. What certainly cannot be explained, as yet, is the mobility and the extent of B-Z junctions as a function of base sequence, but we hope to make some progress in this area soon.

ANGYAN - Do you plan to include the free energy of cavitation in the FIESTA method ?

LAVERY - The inclusion of a cavitation term is certainly one of the possibilities that we are considering. Cavitation effects are clearly not present in the current formulation and may be of importance. The only constraint we impose upon ourselves in making such extensions is that the FIESTA model should remain simple enough for rapid computation in the case of macromolecular systems.

MODELLING IONIC EFFECTS ON BIOMOLECULAR CONFORMATION

D. M. SOUMPASIS^{1,2}

¹Max Planck Institute for Biophysical Chemistry, P.O. Box 2841, 3400
Göttingen (F.R.G.)

²Los Alamos Natl. Lab., Group T-10 M SK-710, Los Alamos, NM 87545
(USA)

SUMMARY

The conformational stability and structural transitions of polyionic biomolecules (e.g. DNA, RNA, polysaccharides, charged proteins etc.) are strongly affected by the interactions with ions. The potential of mean force (PMF) framework has proved to be a computationally feasible, general and often surprisingly accurate technique for the quantitative treatment of their effects caused by biomolecule-diffuse ionic cloud statistical interactions. The main theoretical results obtained so far for DNA-simple electrolyte systems are briefly summarized and discussed below.

INTRODUCTION

An overwhelming amount of experimental results have established that practically all levels of structural organization of highly charged molecules such as the nucleic acids are strongly affected by the types and concentrations of ionic species present in their environment. It is impossible to review this material here but since I will focus on DNA I refer the interested reader to an overview of these phenomena for the case of DNA including many original references. (1).

One of the most dramatic and therefore extensively studied and well characterized ionic effect on DNA structure is the B-Z transition, whereby certain sequences under a variety of environmental conditions undergo a drastic conversion from the usual right-handed form to a completely different left-handed form. The physical chemistry (2) and energetics (3) of this transition have been extensively reviewed, but at this point it is instructive to recall some of the experimental data pertaining to "simple" salt-induced B-Z transitions of linear DNA molecules in the absence of subtleties such as cosolvents, supercoiling etc. Such experimental data sets not only provide valuable insights and

guidance to the theoretician intending to model ion-DNA interactions and their effects on structural stability, but they also provide the necessary benchmarks for assessing the usefulness of his or her theoretical productions.

Table 1 summarizes the critical concentrations (i. e. the salt concentration at which the B and Z form are equiprobable) for the canonical d(C-G) sequence and its modification d(m⁵C-G), whereby the 5 position of C is methylated in various ionic environments.

TABLE 1

Critical concentrations c* of B₁ transitions induced by various salts (chlorides)^a

Sequence	Counterion	c* (mM)	Reference
poly d(C-G)	Na ⁺	2300	(4,5)
	K ⁺	2600	(5)
	Rb ⁺	2900	(5)
	Cs ⁺	4800	(5)
	Mg ²⁺	660	(4)
	^b Zn ²⁺	0.15	(6)
	^b Zn ²⁺ -Cysteine	0.08	(6)
	^b Zn ²⁺ -tris/z amino - ethyl) amine	0.003	(6)
poly d(m ⁵ C-G)	Na ⁺	750	(7)
	Mg ²⁺	0.5	(7)
	^c [Co(NH ₃) ₆] ³⁺	0.02	(7)

a: Transitions at room temperature except for the Zn-complexes (35° C)

b: In the presence of 2mM NaCl

c: In the presence of 50 mM NaCl

It is clear that depending on the ionic species the critical concentrations may vary extremely (a factor of 10⁶ for the case Na⁺ and Zn²⁺-tris-2-aminoethyl amine!) However the alkali-chlorides (and in the case of d(C-G) Mg²⁺ too) induce the transition at high salt concentrations (>1 M) whereas more complex cations (and in the case of d(m⁵C-G), Mg²⁺ too) induce the transition at very low concentrations bearing

stoichiometric relations to DNA bases.

The reason for this behavior is that in the former case the transition is driven by diffuse ionic cloud effects (screening and statistical collisions) whereas in the latter, one has differential stabilization of the Z form due to specific ionic binding to base sites (e. g. N7 of guanine).

These two fundamentally different, general modes of interaction of DNA with ions are competitive (i. e., when both simple and binding cations are present) and are observed with any other charged biomolecule in solution as well.

Modelling of the specific binding mode at the microscopic level requires binding energies (obtained from quantum mechanical calculations), quantitative consideration of subtle hydration-dehydration processes, and treatment of many body effects (since the equilibrium takes place in a dense aqueous charged phase). The outlook for coping with these problems is currently very pessimistic. The diffuse cloud mode is simpler but not simple enough to be microscopically treated using standard available simulation techniques (e. g. Molecular Dynamics, Monte Carlo) primarily due to the combinatorial explosion encountered in applying computer simulations to mixtures (e.g. DNA-water ions), difficulties in obtaining free energies, long range of the Coulomb interaction, etc.

In view of this state of affairs we decided to abandon the currently fashionable philosophy of entirely computer based simulation techniques, adopt cruder models capturing the essential features of the problem only and use semianalytic approximations well embedded in the rigorous theory of statistical mechanics.

The main product of this strategy is a viable approximate framework for quantitative modelling of ionic diffuse cloud contributions to the total conformational free energy of highly charged biomolecular structures (8-12) called the PMF Approach and recently reviewed in Ref. 13.

PMF TREATMENT OF IONIC EFFECTS

We currently use the simplest example of a PMF approach. The central idea involved is the replacement of the bare Coulomb interactions of charged solvent accessible sites on a biomolecule (e.g. phosphates and electronegative base atoms in the case of DNA) by pairwise additive effective interactions (the pmf's) obtained after canonical averaging of all other degrees of freedom (i.e. water and other ions).

In general such pmf's are free energies, i.e. depend on the thermodynamic state parameter of the system envisaged (e.g. salt concentrations, temperature etc.). They make it possible to approximately represent solvent contributions to conformational stability in a very computationally convenient and transferable manner, irrespective of the geometrical complexity of the biomolecule. The model currently adopted for the solvent is the well known Restricted Primitive Model (RPM) of electrolytes (14) picturing water as a dielectric continuum and the hydrated ions as charged hard spheres with distance of closest approach σ .

The RPM-pmf's currently in use are obtained from statistical mechanical approximations such as the semi-analytic Exponential Mean Spherical Approximation (EXP - MSA) (15,16) and the Hypernetted-Chain Approximation (HNC) which involves numerical solution of non-linear integral equations and yields very good results for charged liquids (17).

(More details and references can be found in Refs.8,9,11,12).

Implementation of the approach involves specification of the one parameter σ which in all our work has been fixed at 4.90 Å (see Refs. 8,9). As explained in Ref. (5), the σ 's of other alkali chlorides can be obtained without further adjustment from colligative data of their aqueous solutions. This yields the set of σ 's to be used in RPM-PMF computations: 4.66 Å (KCl), 4.41 Å (RbCl) and 3.77 Å (CsCl).

The remaining intramolecular energy contributions (e.g. covalent bonding, torsional energies, van der Waals interactions, etc.), which to very good approximation do not depend on diffuse ionic cloud interactions, can be represented with semi-empirical force fields routinely used to model biomolecules.

In recent work (11,12) we have used the AMBER force field developed by the USCF group (18).

RESULTS FOR DNA

Using the RPM-PMF outlined above and considering just the phosphate - phosphate interactions suffices to describe quantitatively not only the high salt (>1.5 M NaCl) (8,9) but also the low salt (0.1 - 1.5 M NaCl) (10) dependence of the canonical B-Z transition of d(C-G) helices. With this sequence, inclusion of all other salt independent intramolecular interactions (11,12) does not change the results in any significant way due to cancellations in the free energy balance governing the B-Z isomerization.

This is not true with other sequences such as d(G-C)₁₂ or d(A-T)₆

where our AMBER-PMF force field predicts that a NaCl induced transition is not possible due to too big intramolecular energy differences (between the B and Z conformations) (work in preparation) in accordance with the experimental findings.

Further, the RPM-PMF computation for B-Z transitions of poly d(C-G) induced by the other alkali chlorides (5) reproduces excellently the experimental data of ref. (5) (shown in table 1).

Computation of the full harmonic spectrum and normal modes of d(C-G)₃ in B and Z forms in a wide range of NaCl concentrations (10.01 - 5.0 M), using an AMBER-PMF force field (11) shows that the low frequency vibrations depend on salt concentration and the lowest frequency mode of the B form drastically softens precisely in the salt regime where the transition takes place. This behavior suggests a soft mode mechanism for the B-Z conversion analogous to mode softening in solid state physics (e.g. ferroelectrics).

A study of the relative stabilities of the A,B,C,Z_I,Z_{II} and alternating B conformations considering phosphate - phosphate interactions only (9) has predicted in addition to the B-Z_I case a salt induced B-A transition at 1.85 M NaCl and a B-Z_{II} transition at 0.2 M NaCl.

The latter is not possible (with d(C-G) helices) when one uses the full AMBER-PMF due to large intramolecular energy contributions, but the B-A transition is again obtained for d(G-C)₁₂.

Other applications of the PMF approach to hairpin - B duplex transitions (19), all B to mixed B/Z transitions (20) and helix - coil transitions (21) have been discussed in Ref. (13).

In addition, the approach can be used to approximately determine the 3-dim ionic distributions around DNA (22).

CONCLUDING REMARKS

Real progress in realistic modelling of highly charged biomolecules in solution including their subtle conformational transitions and interactions, is only possible if we somehow manage to handle the formidable complexity of aqueous, multicomponent, inhomogeneous, charged fluids.

Practically none of the results reported in this work could have been obtained using presently widespread simulation techniques (Monte Carlo, Molecular Dynamics) in spite of the availability of supercomputers.

The brute force approach is simply not brute enough to cope with the very brute problems we are interested in.

However, we have provided some evidence that one does not have to keep track of every single particle in the world, in order to obtain approximate albeit very useful answers to relevant questions in biomolecular modelling.

Every many body system, no matter how complex, has to obey the laws of large numbers and the averaging taking place wipes out all but the most relevant facets of the truly microscopic description. Any approximate theory firmly based on statistical mechanics and taking into consideration these most relevant features of the problem is likely to be a useful one and can be often refined step by step.

The simplest version of the PMF Approach based on the RPM description of the solvent and the Kirkwood Superposition Approximation (KSA) is currently being adapted to the treatment of ionic mixtures (for application to specific ionic binding problems) and the KSA is being refined through inclusion of higher order correlations.

At the solvent modelling level, we are testing some structurally more realistic descriptions and will present first results in the near future.

ACKNOWLEDGEMENTS

I wish to thank Drs. A. Garcia, R. Klement, C.S. Tung, E. von Kitzing, T. Jovin and G. Bell for many discussions and collaborations, I am indebted to the U.S. Dept. of Energy and the Bundesministerium fuer Forschung und Technologie FRG for funding.

REFERENCES

- 1 D. M. Soumpasis, in: R. H. Sarma and M. H. Sarma (eds.), *Biomolecular Stereodynamics IV*, Adenine Press, Guilderland N.Y., 1986, pp 47-62
- 2 T. M. Jovin, D. M. Soumpasis and L. P. McIntosh, *Ann. Rev. Phys. Chem* 38, (1987) 521-560
- 3 D. M. Soumpasis and T. M. Jovin, in: F. Eckstein and D. M. J. Lilley (eds.): *Nucleic Acids and Molecular Biology*. Vol. 1. Springer Verlag, Berlin-Heidelberg 1987, pp 85-111
- 4 F. M. Pohl and T. M. Jovin, *J. Mol. Biol.* 67, (1972), 375-396
- 5 D. M. Soumpasis, M. Robert-Nicoud and T. M. Jovin, *FEBS Letters*, 213 (1987), 341-343
- 6 G. V. Fazakerley, *Nucleic Acids Res.*, 12, (1984), 3643 -3649
- 7 M. Behe and G. Felsenfeld, *Proc. Natl. Acad. Sci. USA*, 78, (1981), 1619-1623
- 8 D. M. Soumpasis, *Proc. Natl. Acad. Sci. USA* 81, (1984), 5116-5120
- 9 D. M. Soumpasis, J. Wiechen and T. M. Jovin, *J. Bim. Struct. and Dyn.* 4, (1987), 535-552

- 10 D. M. Soumpasis, *J. Biom. Struct. and Dyn.* **6**, (1988), 563-574
- 11 A. E. Garcia and D. M. Soumpasis, *Proc. Natl. Acad. Sci. USA*, **86**, (1989), 3160-3164
- 12 R. Klement, D. M. Soumpasis, E. von Kitzing and T. Jovin, *Biopolymers*, (1989) (in press)
- 13 D. M. Soumpasis, A. E. Garcia, R. Klement and T. Jovin, in: D. L. Beveridge and R. Lavery (Eds) *Theoretical Biochemistry and Molecular Biophysics*, Adenine Press, Guilderland NY (in press)
- 14 H. J. Friedman and W. D. Dale, in: B. J. Berne (Ed.), *Statistical Mechanics A: Equilibrium Techniques*, Plenum, NY, 1977, pp 85-134
- 15 H. C. Andersen and D. Chandler, *J. Chem. Phys.*, **57**, (1972), 1918-1939
- 16 H. C. Andersen, D. Chandler and J. D. Weeks, *J. Chem. Phys.*, **57**, (1972), 2626-2631
- 17 B. Hafskjold and G. Stell, in: E. W. Montroll and J. Lebowitz (Eds.) *The Liquid State of Matter: Fluids, simple and complex*, North Holland, Amsterdam, 1982, pp. 175-274
- 18 S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta and P. Weiner, *J. Am. Chem. Soc.* **106**, (1984), 765/784
- 19 A. E. Garcia, G. Gupta, C. S. Tung and D. M. Soumpasis, *J. Biom. Struct. and Dyn.*, (1989) (submitted)
- 20 A. E. Garcia and G. Gupta (in preparation)
- 21 A. E. Garcia and D. M. Soumpasis, *Biophys. J.*, **53** (1988), 439 (abstract and in preparation)
- 22 R. Klement, D. M. Soumpasis, T. M. Jovin, *Proc. Natl. Acad. Sci. USA*, (1989) (submitted)

DISCUSSION

PULLMAN - 1) When you discuss the ionic effects or conformations you take into consideration the phosphate backbone of DNA only. While this could be a good approximation for Na^+ , Mg^{++} etc, one should be more careful about Cs^+ because we know from the experimental results of Skuratowski that this ion sits in the grooves of DNA essentially and not on the phosphates. I know why it does so but this is a secondary problem.

2) You indicate that Poly(dA). Poly(dT) is stable only in the B form. This could induce some people to consider this polymer as a typical representative of misform, which it is not. Poly(dA).Poly(dT) has a very particular structure which sets it somewhat apart.

SQUMPASIS - 1) I think that the well localized chain of Cs^+ ions Skuratowski et al have reported in their X-rays DNA fiber studies will not survive the process of dissolution. If it were present in solution it would give rise to physicochemical behavior completely different from the other alkali cations which is not the case as far as I know.

2) It is true that Poly[(dA.dT)] is a special member of the B family not identical to average B. I just wanted to emphasize the fact that one cannot drive it away from the B family under conditions where other members become A, Z, etc.

LAVERY - There are two very interesting cases where it would be useful to know about ionic stabilization effects :

i) curved DNA

ii) DNA with one face neutralized to model protein binding

Can PMF theory be extended to model these asymmetric situations ?

SQUMPASIS - Yes, provided you know the positions of the charges PMF theory will always yield an estimate of the stabilization for energy due to the diffuse ionic cloud irrespective of the geometrical complexity of the structures involved.

WIPFF - Concerning the induced $B \rightarrow Z$ conversion of d(C-G) sequences, you show that Zn (2 aminoethyl) amines complex is very efficient for $\approx 3 \text{ mM}$.

How does that complex interacts with DNA ?

SOUMPASIS - It interacts via specific binding of Zn presumably to NZ of guanine. Further stabilization by H-bonding to nearby DNA atoms is also very likely.

DYMEK - You have experimental data on critical concentrations of sodium and magnesium for B-Z transition, but only for sodium is the free energy difference calculated. Why is it not done for magnesium ?

SOUMPASIS - We have not assigned a distance of closest approach σ for MgCl_2 yet and in addition detailed free energy experimental data were not available until very recently. The Mg case will be treated in a future publication. If one assumes $\sigma = 5.0 \text{ \AA}$ one obtains a transition at 1.0 M MgCl_2 (experimental value 0.66 M MgCl_2) (D.M. Soumpasis, Proc. Natl. Acad. Sci. USA 81, 5116-5120 (1984).

HUIGE - To what extent is the precise form of the PMF potential important for calculations the B-Z transition ?

SOUMPASIS - It is absolutely essential to describe the many body effects of short distance repulsion (e.g. hard spheres within the RPM) as well as possible. Treatment of the Coulomb interactions (e.g. ions modelled as points) alone does not yield a B-Z transition in the experimental regime of monovalent salt concentrations ($> 1\text{M}$).

RULLMANN - To what extent is the PMF method applicable to charge-charge interactions in general ?

SOUMPASIS - The general philosophy of the PFM method applies to any charge-charge interaction in a many body system whatever. The pair PMF is simply the effective pair interaction of any two charges at given fixed positions when all other degrees of freedom have been statistically averaged. However, the accuracy of the number σ you get will of course depend on both the structural model of the system and the approximations you use to do the averaging.

606

OHLENBUSCH - Would you be able to take the coordinates of my chromatin subunit model to make estimates about its stability.

SQUMPASIS - Yes.

INFLUENCE OF THE BASE SEQUENCE AND CONFORMATION ON THE STRUCTURE
OF THE POLYNUCLEOTIDE HYDRATION SHELL

V.I.POLTEV¹, A.V.TEPLUKHIN² and G.G.MALENKOV³

¹Institute of Biological Physics, USSR Academy of Sciences,
Pushchino, Moscow Region, 142292, USSR

²Computing Research Center, USSR Academy of Sciences, Pushchino,
Moscow Region, 142292, USSR

³Institute of Physical Chemistry, USSR Academy of Sciences,
Moscow, Leninsky Prospekt, 31, 117915, Moscow USSR

SUMMARY

Hydration shells in major and minor grooves of double helices with various nucleotide sequence and conformation have been studied. Monte-Carlo simulation of the systems containing a stack of complementary base pairs and 30 water molecules per pair has been performed. Characteristic features of the hydration shell structure has been found for the stacks of repeating A:T and G:C base pairs as well as alternating (A:T, T:A) and (G:C, C:G) ones. Presence of common features in arrangement of hydrophilic centres common for various sequences and configurations is manifested in the similar structural elements of the hydration shells. Probabilities of the formation of bridges, formed by 1, 2 and 3 water molecules, between hydrophilic centres of the bases have been estimated. Hydration shell structure was found to depend significantly on the stack sequence and configuration, while global hydration characteristics are only slightly dependent on the nature of the stack. For the stacks in A configuration the number of water molecules forming more than one H-bonds with the bases is greater in comparison with the stacks in B-like configuration. This result is discussed in connection with the concept of hydration economy during B to A transition.

INTRODUCTION

The mutual influence of the space organization of the polynucleotide double helix and the structure of its environment on each other can be considered as determined experimentally. One of the most important manifestations of the properties of the polynucleotide fragments with different sequence and conformation is difference in their hydration. Thus, double layer water spine is characteristic for the central part of d(CGCGAATTCGCG) dodecamer (ref. 1), while in the case the d(CCAAGATTGG) decamer crystal (ref. 2) and that of modified d(GCGGCG) hexamer (ref. 3) water bridges ("springs") between N3 or O2 base atoms and sugar ring oxygen of the neighbor nucleotide are typical elements of the hydration shell structure. These differences can be reflected in specific interactions of DNA with proteins and other molecules of biological importance. Regular construction from water molecules in the vicinity of DNA surface can be replaced by the other molecules. Thus netropsin can, according to X-ray data, replace double layer water spine in

minor groove of double stranded dodecamer d(CGCGAATTCGCG) (ref. 4).

Interactions with water molecules can give rise to conformational peculiarities in certain parts of DNA, which are important for biological functioning of the biopolymer. Computer simulation is one of the most powerful tools in the study of biopolymer-water interaction. This simulation can provide us information which can not be obtained by conventional experimental methods. The study of polynucleotide - water systems by computer simulation are successfully performed during last decade in several laboratories and the interesting results on hydration shell structure of DNA, its components and complexes have been already obtained (ref. 5-10).

Our group is engaged in a systematic investigation of the hydration of nucleic acid fragments using Monte-Carlo simulation. Earlier we have developed methods of calculating energy in related systems (ref. 7), simulated hydration of bases and complementary pairs (ref. 8), as well as performed the simulation of hydration shells of minor and major grooves of poly(dA) poly(dT) in the two different B conformations (ref. 10).

Here we present some results of simulation of the hydration of the stacks formed by complementary base pairs with different nucleotide sequences: $(A:U)_n$, $(G:C)_n$, $(A:U,U:A)_n$, $(G:C,C:G)_n$. We consider the dependence in hydration of certain hydrophilic centres on the stack configuration and sequence, determine probability of water bridges between base atoms, elucidate and specify the most typical fragments of the hydration shell. The simulations showed that subtle peculiarities of hydration shell structure but not the global characteristics strongly depend on sequence and configuration of the stack. Some preliminary results of this work was included in early publication (ref. 9).

METHODS OF CALCULATIONS

All the simulated systems contained a stack of six base pairs arranged as in a certain conformation of double helix and 180 water molecules. N1 pyrimidine and N9 purine atoms were methylated. Helical periodic boundary conditions were imposed along the helical axis at the stack. Distance between any water molecule and centre of at least of one of the bases was not permitted to be greater than 9.5Å.

Mutual arrangement of the bases which in general features resemble those found in A- and B-forms of DNA were considered for each sequence. For repeating AU stack additional B'-conformation resembling that in the central part of d(CGCGAATTCGCG) dodecamer (ref. 1) was considered. For other sequences such base arrangement are extremely unfavorable energetically.

Specially devised for such systems potential functions (ref. 7,12) have been used for calculating interaction energy. 1-6-12 type potential func-

tions are used to calculate interactions between all the atoms, except those, involved in H-bonding. In the latter case 1-10-12 functions are used. Geometrical criterion was used for the determination of the H-bonds: A-H...B configuration was called an H-bond if A...B distance was less than 3.2Å and H...B distance was less than 2.4Å, where A and B are O or N atoms.

Average energetical and structural hydration shell characteristics were calculated from a statistically significant sampling of configurations obtained by a Metropolis et al. algorithm (ref. 13). Each of the states, accepted by a Metropolis procedure corresponds to an instantaneous (I) structure of the hydration shell. Along with analyses of the ensemble of the I structures we obtained "frozen" (F) structures (14) by lowering the temperature of simulation up to 5K. Positions of water molecules in F structures are close to those, corresponding to local potential energy minima. For each system studied number of obtained I structures was about $2 \cdot 10^6$.

RESULTS AND DISCUSSION

General regularities.

Total characteristics of the hydration shells of the stacks are shown in the Table 1.

TABLE 1

Energy and structure characteristics of hydration shells of base pair stacks

sequence	(A:U) _n			(A:U,U:A) _n		(G:C,C:G) _n		(G:C) _n	
configuration	A	B	B'	A	B	A	B	A	B
energy values ^a									
total	-906	-905	-901	-887	-896	-945	-953	-970	-982
water-water	-7.16	-7.26	-7.26	-6.99	-7.20	-6.91	-7.08	-7.03	-7.12
water-pur	-23.0	-22.0	-22.5	-22.9	-22.0	-30.8	-30.2	-34.4	-33.6
water-pyr	-20.7	-19.8	-18.8	-20.1	-19.4	-23.1	-22.4	-21.8	-23.1
total numbers of H-bonds ^a									
water-water	1.75	1.76	1.75	1.67	1.71	1.68	1.70	1.74	1.76
water-pur	3.10	3.05	3.20	3.14	2.82	4.61	4.68	4.37	4.38
water-pyr	2.39	2.38	1.94	2.44	2.16	2.24	1.95	1.79	1.84
numbers of water molecules H-bonded to base atoms ^a									
N7	1.40	1.51	1.59	1.41	1.11	1.60	1.58	1.37	1.36
O4,O6	1.38	1.23	1.20	1.44	1.09	1.04	1.16	1.00	1.06
N6-H,N4-H	0.80	0.91	0.87	0.83	0.73	0.93	0.90	0.68	0.81
N3	0.88	0.63	0.74	0.90	0.97	1.08	0.99	1.07	0.98
O2	1.01	1.15	0.74	0.99	1.07	1.30	1.04	1.07	1.02
N2-H						0.90	0.96	0.92	0.98

^aTotal energy refers to six base pairs and 180 water molecules; the water-water and water-base energy values and H-bond numbers refer to one water molecule and one base respectively.

Average total potential energy of the systems both for A and B stacks are lower for G:C-containing stacks than for A:U ones as it was the case for separate base pair - water systems (ref. 8). Our simulations show, that total hydration characteristics, such as energy and its components, number of water-water and water-base H-bonds only slightly depend on the stack sequence and configuration. At the same time, subtle peculiarities of the hydration shell structure in both grooves of the double helix change essentially when stack configuration and (or) sequence is transformed. This change of hydration shell structure can be most vividly demonstrated if we consider the pattern of water bridges formed between hydrophilic centres of the bases. Common for various sequences and configurations features in arrangement of hydrophilic centres (ref. 11) lead to common structural elements of hydration shells.

Water bridges between hydrophilic centres

Three types of water bridges are taken into consideration: those formed by one (B1), two (B2) and three (B3) water molecules. Situation, corresponding to B1 bridge, that is when a water molecule forms H-bonds with two (sometimes three) hydrophilic centres simultaneously, corresponds almost to all potential energy minima in system base + 1 water molecule (ref. 7). Arrangements, corresponding to B2 bridges (when each of the two water molecules form an H-bond between each other and with a hydrophilic centre) are typical for potential minima in the systems base + 2 water molecules (ref. 7). In B3 bridge two water molecules form H-bond with hydrophilic centres and with a third molecule, which is not H-bonded to bases.

Several water bridges may connect the same pair of hydrophilic centres in some F and I structures. The B1 bridges are formed with high probability if N...N, N...O or O...O distance between the centres is close to 4.3Å. Probability of bridge formation between bases, belonging to the same pair is less in stacks than in separate base pairs (ref. 8). This is due to formation of a space H-bonded network comprising the stack in which many bridges between centres belonging to different base-pairs are formed. Such bridges make contribution to the stabilization of the stack, difference in probabilities of their formation manifesting to different degree of stabilization of various DNA conformation by aqueous environment. Let us consider some water bridges in more detail and describe characteristic elements of the hydration shell structure in the minor groove of the double helix. These elements can be defined when considering F structures (Fig. 1 is an example). Only those water molecules are shown in this figure which participate in formation of H-bonds with the bases or of water bridge.

Water bridges and hydration shell structure in minor groove.

Total probability of B1 bridge formation in this groove for B stacks is rather low (less than 16%). Location of water molecules, forming such

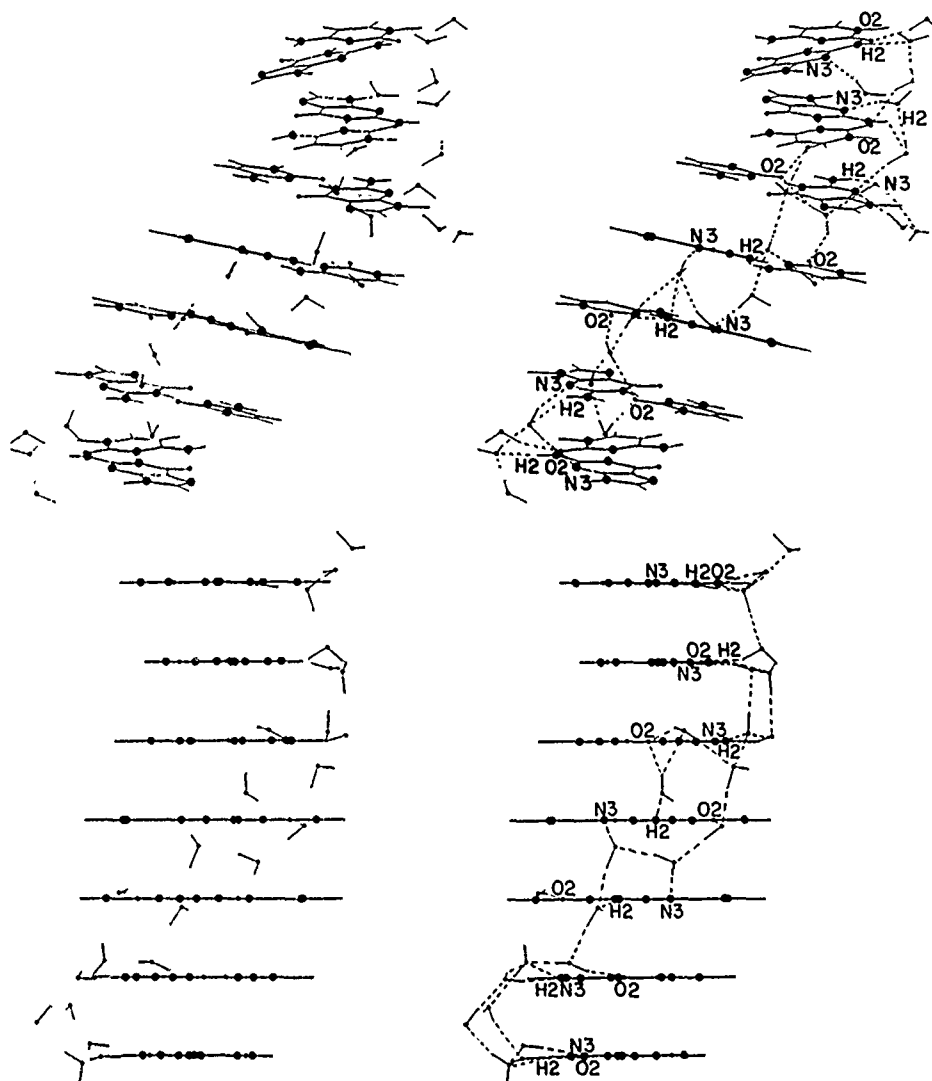
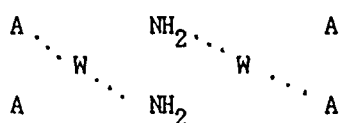


Fig. 1. Water-base hydrogen bonds and water bridges in the minor groove of alternating G:C stack. Stereo view of F-structure of hydration shell for A (top) and B (bottom) configuration.

bridges in A:U stacks is strikingly different from that characteristic for G:C stacks, which is connected with the presence of an additional hydrophilic centre - amino group. Bridges with participation of this group most frequently are formed according to scheme



[illegible]

Hydration economy and B to A transitions

NM/NHB ratio characterizes the degree of "water economy" in the grooves. This ratio is less (water economy is greater) in major groove than in minor one for all the studied stacks. "Hydration economy" concept was proposed by Saenger and co-workers (ref. 15) to explain the mechanism of B to A transition. As it was noted (ref. 15) two oxygens from neighbour phosphate groups can be connected in A form by one water molecule (B1 bridge according to our denotation). Distance between these atoms being greater in B-form, formation

of such a bridge becomes impossible, and at least two water molecules are necessary to hydrate phosphate group of B DNA.

In our model system there are no phosphate groups, but still water economy is observed if we compare structure of hydration shells of stacks in A and B configurations. It is manifested in decrease of NM/NHB at B to A transition. On the F structure level this trend can be seen for all the sequences studied. For I-structure ensemble this is not the case for repeating G:C stack. Ability of poly(dG):poly(dC) to convert easily to A-form can be explained by the interaction of the adjacent G:C base pairs which is more favorable in A-conformation (ref. 11).

TABLE 2

Characteristics of hydration shells of double helix grooves^a

configuration		A			B		
groove		1	2	1+2	1	2	1+2
NHB	(A:U)	11.36	21.11	32.47	10.69	21.86	32.55
	(A:U,U:A)	11.35	22.13	33.48	12.21	17.66	29.87
	(G:C,C:G)	19.65	21.43	41.08	17.93	21.88	39.81
	(G:C)	18.36	18.55	36.91	17.89	19.41	37.30
NW	(A:U)	9.0	14.2	23.2	10.2	17.4	27.6
	(A:U,U:A)	10.4	15.2	25.6	11.6	13.6	25.2
	(G:C,C:G)	16.8	16.4	33.2	16.6	16.1	32.7
	(G:C)	17.3	15.1	32.4	15.5	15.1	30.6
NW/NHB	(A:U)	0.79	0.67	0.72	0.95	0.79	0.85
	(A:U,U:A)	0.92	0.68	0.76	0.95	0.77	0.84
	(G:C,C:G)	0.86	0.76	0.81	0.93	0.74	0.82
	(G:C)	0.94	0.81	0.88	0.87	0.78	0.82

^aNHB - the number of water-base H-bonds, NM - the number of water molecules forming H-bonds with bases. All values refer to six base pairs. Column 1 corresponds to minor groove, 2 - major.

Though our calculations do not take into account difference in A- and B-configurations for various sequences nor hydration of sugar-phosphate backbone, qualitative results correctly reflect the trend of the bases to arrange in such a way, that decrease of water in the system will result in decrease of number of water molecule necessary to hydrate all the hydrophilic centres.

On the possible influence of sugar-phosphate backbone.

For correct description of the DNA-water systems, contribution of sugar phosphate backbone and of counter ions should be also taken into account. Unfortunately, our computer resources are not sufficient to simulate such a system, for it should contain much more water molecules. These

contributions may change pattern of the base hydration in the grooves. But this change should be rather small if the groove is wide enough. This is the case for all configuration considered except the minor groove of B' form. The minor groove of B'-conformation is narrowed, and its pattern of hydration differs significantly for base pair stack from that of B'-DNA helix. But our results help to emphasize the rôle of base arrangement in hydration shell structure in this case as well.

REFERENCES

- 1 R.E. Dickerson, H.R. Drew and B. Conner, Single crystal X-ray structure analysis of the A, B and Z helices or one good turn deserves another, in: R.H. Sarma (Ed.) *Biomolecular Stereodynamics*, Adenine Press. N.Y., 1982, pp.1-34.
- 2 G.G. Prive, U. Heinemann, S.Chandrasegaran, L.-S. Kan, and R.E. Dickerson, Helix geometry, hydration and G A mismatch in a B-DNA decamer, *Science*, 238 (1987) 498-504.
- 3 W.B.T. Cruse, S.A. Salisbury, T. Brown, R. Cosstick, F. Eckstein and O. Kennard, Chiral phosphorothioate analogues of B-DNA. The crystal structure of Rp-d(Gp(S)CpGp(S)CpGp(S)C), *J.Mol.Biol.*, 192 (1986) 891-905.
- 4 M.I. Kopka, Ch. Yoon, D. Goodsell, P. Pjura and R.E. Dickerson, The molecular origin of DNA-drug specificity in netropsin and distamycin, *Proc. Natl. Acad. Sci. USA*, 82 (1985) 1367-1380.
- 5 E. Clementi, Structure of water and counterions for nucleic acids in solution, in: E. Clementi and R.H. Sarma (Eds) *Structure and Dynamics: Nucleic Acids and Proteins*, Adenine Press, New York, 1983, pp.321-364.
- 6 P.S. Subramanian, G. Ravishanker and D.L. Beveridge, Theoretical consideration on the "Spine of hydration" in the minor groove of d(OGGAATTGOGG).d(GGGCTTAAGGGC): Monte-Carlo computer simulation, *Proc. Natl. Acad. Sci. USA*, 85 (1988) 1836-1840.
- 7 V.I. Poltev, T.I. Grokhilina and G.G. Malenkov, Hydration of nucleic acid bases studied using novel atom-atom potential functions, *J. Biomol. Struct. Dyn.*, 2 (1984) 413-429.
- 8 V.I. Poltev, N.V. Shulyupina, A.V. Teplukhin and G.G. Malenkov, Investigation of hydration features of nucleic acid bases and complementary base pairs by the Monte-Carlo method, *Mol. Biol. (USSR)* 21, (1987) 1600-1609, *Engl. Transl. Molecular Biology* 21 (1987) 1107-1113.
- 9 A.V. Teplukhin, V.I. Poltev and G.G. Malenkov, Monte Carlo simulation of the hydration of the nucleic acid fragments, *J. Biomol. Struct. Dyn.* 7 (1989) 75-99.
- 10 V.I. Poltev, A.V. Teplukhin and V.P. Chuprina, Conformational behavior and water shells of different DNA duplexes. Computer simulation biological significance, in: E. Clementi and S. Chin (Eds.), *Biological and Artificial Intelligence Systems*, ESCOM, Leiden, 1988, pp.163-184.
- 11 V.I. Poltev and A.V. Teplukhin, Conformational implications of some nucleotide sequences, *Int. J. Quant. Chem* 35 (1989) 91-102.
- 12 V.I. Poltev and N.V. Shulyupina, Simulation of interaction between nucleic acid bases by refined atom-atom potential functions, *J. Biomol. Struct. Dyn.*, 3 (1986) 739-765.
- 13 N.A. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller and E. Teller, Equation of state calculations by fast computing machines, *J. Chem. Phys.* 21 (1953) 1087-1092.
- 14 G.G. Malenkov, Models for the structure of hydration shells of simple ions based on crystal structure data and computer simulation, in: R. Dagonadze et al. (Eds), *Chemical Physics of Solvation*, Elsevier, Amsterdam, 1986, part A., pp.355-390
- 15 W. Saenger, W.N. Hunter and O. Kennard, DNA conformation is determined by economics in the hydration of phosphate groups, *Nature*, 324 (1986) 385-388.

DISCUSSION

OHLENBUSCH - You have considered DNA to be a straight rod, but in solution DNA is generally bent so that grooves vary in size considerably depending upon whether they are on the inside or outside of the bend. Have you considered this effect in your calculations ?

POLTEV - We have not carried out calculations for bent duplexes as nucleotide sequences considered have no tendency to be bent. Bending takes place on the boundary between different conformations ; we considered only regular helices yet.

MAROUN - You define a "water economy" factor and calculated it for A and B-DNA trying to show support for the water economy idea of Sanger et al. Did you calculate this factor for Z-DNA then ?

POLTEV - No calculations for Z conformation were performed by us till now. I think we can try to estimate the characteristics of base hydration for this form also.

GHOMI - What are the main structural characteristic of the B' conformation, especially for homopurine-homopyrimidine double-helices ?

POLTEV - The main structural characteristic of B' conformation is narrowed minor groove. We performed calculations for configuration with negative tilt ($\sim 10^\circ$) and nearly zero propeller twist. Other low energy B' conformations can have large propeller twist and nearly zero tilt. We hope the main hydration features for various B' configurations will be rather similar.

WIPFE - Concerning the polymorphism of DNA as a function of humidity : changing the humidity charges also the concentration of cations near the phosphate in the groove. What qualitative feature in hydration pattern would you expect from the explicit consideration of these ions in the simulation ? Cation obviously give stronger interactions with water than the bases.

POLTEV - The explicit consideration of ions in the simulation will result in changing the global but not the relative, for various sequences, characteristics of duplex hydration. We hope that the differences in hydration patterns for various sequences will be similar to those obtained without taking ions into consideration (except the helices with narrowed groove).

THEORETICAL PREDICTION OF BASE PAIR SEQUENCE EFFECTS IN Z-DNA AND Z-RNA

B. HARTMANN and R. LAVERY

Institut de Biologie Physico-Chimique, 13 rue Pierre et Marie
Curie, Paris 75005 (France)

SUMMARY

Base sequence effects within double stranded DNA and RNA oligomers in the Z conformation have been studied by molecular modeling using a methodological approach specifically adapted to nucleic acids. Calculations on symmetric oligomers having homonucleotide or dinucleotide repeating base sequences show that sequence changes can produce modifications in overall conformation and strongly affect stability. Within the Z family, it is demonstrated that certain sequences can adopt more than one polymorphic form. Enthalpies of transition, B-Z for DNA and A-Z for RNA, are calculated as a function of base sequence.

INTRODUCTION

In contrast to the situation for B-DNA, although the role of left-handed Z-DNA has been implicated in several biological systems (1,2,3), little is known about the influence of base sequence for this conformation other than the fact that such effects exist (4,5,6). Likewise, it has been recently confirmed that double stranded RNA can exist in the Z form, but only a single crystal $\text{Br}_{\text{GC}}\text{Br}_{\text{GC}}$ showing this conformation has been reported (7).

These facts have encouraged us to undertake a general molecular modeling project aimed at understanding the conformational and energetic details of Z forms of DNA and RNA as a function of base sequence.

METHODOLOGY

The calculations presented have been performed with the Jumna algorithm which was specifically designed for energy minimisation of nucleic acid oligomers. This methodology and the energy formulation are fully described elsewhere (8,9).

All the calculations reported refer to 8 or 12 base pair oligomers of DNA and RNA, respectively. We have studied Z

conformations with all possible homonucleotide and dinucleotide repeating sequences. In consequence, we have been able to impose dinucleotide symmetry on each oligomer treated. Exact symmetry is maintained within each strand, but heteronomous differences are allowed to develop between the strands of the duplex. We calculate the "environmental" energy (E_{env}) of the central dinucleotide as follows,

$$E_{env} = E_c + 1/2 \sum_j E_{cj},$$

where E_c is the internal energy of the central dinucleotide pair and E_{cj} is the interaction energy between this pair and the other dinucleotide pairs forming the oligomer. This has the advantage of both reducing the computational time for energy minimisation and also of eliminating end effects.

In order to represent the electrostatic damping associated with aqueous solution and the presence of a counterion atmosphere, we reduce the net charge on each phosphate group to -0.5 and we employ a distance dependent dielectric function (10). This function has a slope $S=0.356$ corresponding to strong damping conditions (Z conformation).

Due to the alternation of syn and anti nucleotides in the backbones of Z conformation, 10 oligomers are necessary to cover all the unique homonucleotide and dinucleotide sequences. The first base of these dimers is always a syn nucleotide while the second is always anti. The number of errors listed refers to the number of times the usual correlation syn-purine/anti-pyrimidine is violated for each dimer. (we only show one strand of the duplex concerned, in the 5'-3' sense) :

```

..CGCGCGCG.. : G(syn)pC(anti) : 0 error
..CACACACA.. : A(syn)pC(anti) : 0 error ≡ GpT or GpU
..TATATATA.. : A(syn)pT(anti) : 0 error
..UAUAUAUA.. : A(syn)pU(anti) : 0 error
..GGGGGGGG.. : G(syn)pG(anti) : 1 error ≡ CpC
..AAAAAAAA.. : A(syn)pA(anti) : 1 error ≡ TpT or UpU
..GAGAGAGA.. : A(syn)pG(anti) : 1 error ≡ CpT or CpU
..AGAGAGAG.. : G(syn)pA(anti) : 1 error ≡ TpC or UpC
..GCGCGCGC.. : C(syn)pG(anti) : 2 errors
..ACACACAC.. : C(syn)pA(anti) : 2 errors ≡ UpG or UpG
..ATATATAT.. : T(syn)pA(anti) : 2 errors
..AUUAUAUA.. : U(syn)pA(anti) : 2 errors

```

From now on, following the central dimer, these fragments are referred to respectively as: GC, AC, AT, AU, GG, AA, AG, GA, CG, CA, TA and UA.

As starting conformations for energy minimisation we have used the idealized Z_I crystallographic results of Rich (11). However, changes in starting geometry were found to be unimportant. In particular, modification of the sequence of an energy optimised oligomer, followed by reminimisation, allowed all the conformational features of the new sequence to be recovered.

RESULTS AND DISCUSSION

The energy optimisation of the oligomers described above leads to more the results show in table 1. Note that for certain base sequence more than one possible low energy conformation was found. In addition, considerable changes in conformation were observed which can be characterized by the shift of the base pairs from the helical axis. In the table the various results are segregated on the basis of this parameter.

TABLE 1
Energies (E_{env}) (kcal/mol) of Z-DNA and Z-RNA as a function of sequence, ordered horizontally in terms of base pair shift.

DNA Sequence	$\leq 3 \text{ \AA}$	$> 3 \text{ \AA}$	RNA Sequence	$\leq 3 \text{ \AA}$	$> 3 \text{ \AA}$
GC	- 22.3		GC	7.9	17.2
AC	- 13.2		AC	15.0	22.2
AT	- 6.7		AU	21.5	33.1
GG	- 19.1	- 19.9	GG	10.6	11.9
AA	- 4.2	- 3.8	AA	23.5	28.0
AG	- 11.2	- 11.8	AG	15.9	20.5
GA	- 12.6	- 12.8	GA	17.8	19.2
CG		- 18.7	CG	12.0	12.4
CA		- 11.2	CA	19.7	20.1
TA	- 3.3	- 3.2	UA	24.9	27.6

If we consider the lowest energy state in each case, we obtain the following order of stability:

For Z-DNA:

Dimers: GC > GG > CG > AC > GA > AG > CA > AT > AA > TA
Errors: 0 1 2 0 1 1 2 0 1 2

For Z-RNA:

Dimers: GC > GG > CG > AC > AG > GA > CA > AU > AA > UA
Errors: 0 1 2 0 1 1 2 0 1 2

It can be seen from this order that the classical G(syn)pC(anti) sequence is effectively found to be the most stable in the Z form of both nucleic acids. The energetic order seen primarily correlates with the number of GC pairs in each sequence, but, within this division, there is also a clear ordering in terms of the number of errors.

In order to estimate B-Z or A-Z transition enthalpies we have also studied B or A oligomers under high dielectric damping conditions (table 2). These results have been described in detail in previous publications (12,13)

TABLE 2
Energies (E_{env}) (kcal/mol). of B-DNA and A-RNA as a function of sequence under high dielectric damping conditions

DNA Sequence	Energy	RNA Sequence	Energy
GC	- 26.8	GC	- 3.2
AC	- 18.8	AC	5.5
AT	- 11.9	AU	12.3
GG	- 26.8	GG	- 1.9
AA	- 13.1	AA	10.3
GA	- 19.5	GA	5.2

From these results we can now estimate the enthalpies of transition (ΔH tran) between B or A and Z allomorphs. Although the present calculations include neither entropic effects nor explicit solvent or counterion interactions, it is interesting to compare the order we deduce with available experimental results for Z-DNA. These results, transition free energies, have been directly obtained experimentally only for the sequences indicated by a star, the remaining values being deduced by simple arguments based on stacking energies (14).

B-Z transition for DNA:

GC > AT > AC > GA > GG > AG > CA = TA > CG > AA
 ΔH tran: 4.5 4.9 5.2 6.7 6.9 7.6 7.7 8.1 8.6 8.9
 Errors: 0 0 0 1 1 1 2 2 2 1
 Experimental: GC*> AC*> AT*= GG*> GA*> AG > AA > CG > CA > TA

A-Z transition for RNA:

AU > AC > GC = GA > GG = AG = UA > AA > CG = CA
 ΔH tran: 9.2 10.5 10.8 10.8 12.5 12.6 12.6 13.1 15.2 15.3
 Errors: 0 0 0 1 1 1 2 1 2 2

For Z-DNA, we note a good correlation between experimental and predicted orders. We can see from all these results that there is a strong correlation with the number of syn-purine/anti-pyrimidine errors. It thus appears that the relative instability of pyrimidine nucleotides in the syn conformation, which has been recognised both experimentally and theoretically (15 and references therein), is an important factor controlling the stability of the Z conformation of sequence.

One can also note from these results that the relative difficulty of converting RNA to the Z conformation (16) is reflected by calculated transition enthalpies which are roughly twice as large as those we have obtained for DNA oligomers.

For many of the sequences studied we have been able to locate several stable conformations. The details of all these structures are given in tables 3 and 4. The most appropriate way to distinguish amongst the different conformational possibilities is the shift of the base pairs from the helical axis since very large variations of this parameter are found. It can be seen that the shift is related to the number of errors : no errors implies small shifts, one error can lead to either small or large shifts and two errors can generally lead to large shifts (see also table 1).

We now return to the details of the conformations. We note a strong differentiation of the twist between the syn-p-anti steps and the anti-p-syn steps. Generally, a small twist (anti-p-syn step) corresponds to a large rise and vice versa.

For small shift conformations, we can also note several common moderate positive tilts and small negative propeller twists. The Z conformation of RNA leads to the possibility of forming a variety of sugar-base hydrogen bonds in the minor groove. This possibility is limited to syn-p-anti dinucleotide steps where the large twist brings the ribose hydroxyl of the anti-nucleotide close to the base of the anti-nucleotide in the opposing strand. The existence of such bonding, which occurs whenever the latter base is a pyrimidine (implying bonding to the O2 atom) or adenine (bonding to N3), is a contributing factor to the small shift values and the bucking outwards of both base pairs forming the syn-p-anti dinucleotide.

Turning to the structures with large shift, we can note moderate propeller twists but large negative tilts, and an overall variation in twists which increase for the syn-p-anti steps and decrease for the the anti-p-syn steps. For the Z-RNA intra or

TABLE 3
Helicoidal parameters for the optimized conformations of Z-DNA and Z-RNA as a function of base sequence.

Z-DNA					Z-RNA				
Seq.	Shift	Tilt	Prop.	Buckle	Seq.	Shift	Tilt	Prop.	Buckle
Small shift									
GC	1.2	14.5	-1.1	4.9	GC	0.5	10.6	- 7.9	-11.1
	1.2	14.5	-1.1	-4.9		0.5	10.6	- 7.9	11.1
AC	1.5	10.3	0.5	8.6	AC	0.6	8.6	-10.5	-11.7
	1.5	10.7	-1.4	-4.4		0.7	9.1	-11.8	10.9
AT	1.0	4.6	-3.6	0.3	AU	0.5	7.8	- 8.4	-11.6
	1.0	4.6	-3.6	-0.3		0.5	7.8	- 8.4	11.6
GG	1.1	14.4	-0.3	5.0	GG	0.1	14.8	- 6.2	- 5.0
	1.3	12.0	1.8	-4.9		0.3	14.3	- 3.1	13.5
AA	2.3	6.0	4.2	8.8	AA	0.9	4.8	- 3.7	- 7.0
	2.3	6.2	-3.4	-10.9		1.0	5.2	- 3.2	8.8
AG	1.8	8.3	0.7	12.2	AG	0.1	12.6	- 1.6	3.9
	1.7	10.4	5.8	-11.9		0.4	16.7	-10.1	-16.5
GA	3.0	2.2	-7.6	12.8	GA	0.3	10.5	- 8.8	-10.3
	3.1	4.0	-1.5	5.5		0.4	8.8	- 2.7	-12.4
					CG	2.1	8.7	- 3.4	5.8
						2.1	8.7	- 3.4	- 5.8
					CA	2.1	8.1	- 1.8	1.8
						2.0	8.7	- 4.6	- 5.6
TA	1.5	10.0	-5.7	2.0	UA	1.5	4.3	- 8.1	- 8.8
	1.5	10.0	5.7	-2.0		1.5	4.3	- 8.1	8.8
Large shift									
					GC	5.9	-21.0	- 2.9	7.8
						5.9	-21.0	- 2.9	- 7.8
					AC	6.0	-24.0	2.1	11.4
						5.8	-19.1	- 7.4	-12.2
					AU	5.3	-21.0	- 0.4	10.7
						5.3	-21.0	- 0.4	-10.7
GG	4.8	- 3.1	-7.2	13.3	GG	5.3	- 9.8	-12.0	9.4
	4.9	-10.3	0.5	-14.8		5.4	-17.1	0.3	-15.7
AA	6.1	-21.5	0.9	15.0	AA	5.5	-18.1	- 1.4	13.7
	6.0	-23.1	-1.5	-17.6		5.5	-19.0	- 2.0	-12.8
AG	6.6	-22.5	4.9	20.1	AG	5.5	-20.6	2.8	-10.2
	6.5	-18.6	0.1	-20.4		5.3	-18.7	- 3.6	11.8
GA	4.8	-12.3	-3.1	17.0	GA	5.5	- 9.9	-12.9	-12.0
	4.8	- 5.2	-8.4	-11.1		5.5	-17.6	- 3.2	17.3
CG	5.4	-13.3	4.1	13.7	CG	5.4	-10.6	4.5	15.9
	5.4	-13.3	4.1	13.7		5.4	-10.6	4.5	-15.9
CA	4.8	-10.2	3.0	11.6	CA	4.8	- 8.3	4.2	13.6
	4.8	- 9.0	0.6	-11.8		4.8	- 5.8	0.3	-12.5
TA	4.5	-10.4	0.7	8.8	UA	4.7	-10.0	0.1	9.7
	4.5	-10.4	0.7	8.8		4.7	-10.0	0.1	- 9.7

TABLE 4
Helicoidal parameters obtained for the different Z-RNA oligomers
(all anti-nucleotides are indicated by a quote).

Z-DNA				Z-RNA			
Sequence	Dimer	Rise	Twist	Sequence	Dimer	Rise	Twist
Small shift							
GC'	C'pG	4.7	- 8.9	GC'	C'pG	3.6	-13.2
	GpC'	2.7	-49.2		GpC'	3.3	-50.1
AC'	C'pA	4.6	-10.0	AC'	C'pA	3.5	-14.3
	ApC'	2.8	-46.1		ApC'	3.4	-51.1
AT'	T'pA	3.7	-14.6	AU'	U'pA	3.3	-14.9
	ApT'	3.0	-41.6		ApU'	3.4	-49.7
GG'	G'pG	4.7	-11.0	GG'	G'pG	3.9	-14.4
	GpG'	2.6	-45.0		GpG'	2.9	-45.7
AA'	A'pA	4.4	-10.5	AA'	A'pA	3.4	-13.6
	ApA'	2.7	-44.4		ApA'	3.3	-45.5
GA'	A'pG	4.2	- 8.1	GA'	A'pG	3.9	-14.9
	GpA'	2.8	-50.5		GpA'	3.0	-45.2
AG'	G'pA	5.0	-11.7	AG'	G'pA	3.6	-14.0
	ApG'	2.4	-39.9		ApG'	3.2	-45.1
TA'	A'pT	4.4	- 9.9	CG'	G'pC	4.7	-10.4
					CpG'	2.5	-46.8
				CA'	A'pC	4.4	-10.3
					CpA'	2.7	-48.4
				UA'	A'pU	3.4	-14.0
					UpA'	3.2	-46.3
Large shift							
GG'	G'pG	3.7	- 8.0	GC'	C'pG	1.7	- 4.5
					GpC'	3.5	-63.7
				AC'	C'pA	1.9	- 5.9
					ApC'	3.2	-61.6
				AU'	U'pA	2.1	-10.3
					ApU'	3.4	-54.7
GG'	G'pG	3.0	- 8.2				
	GpG'	2.9	-53.9				
AA'	A'pA	2.0	- 7.2	AA'	A'pA	2.3	- 7.7
	ApA'	3.1	-57.6		ApA'	3.1	-57.1
AG'	G'pA	3.4	- 6.9	AG'	G'pA	2.2	- 9.8
	ApG'	2.7	-52.9		ApG'	3.3	-54.8
GA'	A'pG	2.5	- 6.7	GA'	A'pG	3.0	- 5.5
	GpA'	2.8	-53.4		GpA'	2.7	-56.8
CG'	G'pC	3.3	- 8.5	CG'	G'pC	3.7	- 7.5
	CpG'	2.9	-48.8		CpG'	2.8	-48.7
CA'	A'pC	3.4	- 7.7	CA'	A'pC	3.8	- 7.1
	CpA'	2.9	-49.8		CpA'	2.8	-50.1
TA'	A'pT	3.1	- 7.4	UA'	A'pU	3.2	- 6.6
	TpA'	3.1	-51.8		UpA'	3.0	-52.9

inter sugar-base hydrogen bonds can be formed : between the guanine amino group and a sugar O2' in the opposing strand (for GG and AG), between the H(N2) atom of guanine and the O2' ribose hydroxyl of the 3' side (for GC and AC) or between the N3 atoms of guanine or adenine and the ribose hydroxyl of the attached sugar (for GG, CG and CA).

In all cases, we find that syn nucleotides are associated with C3'-endo sugars, while anti nucleotides have C2'-endo sugars. Despite the important changes in helicoidal parameters we have described, the backbone geometry remains remarkably close to those of the Z_I conformation, and presents very little variation.

For Z-DNA, the most stable calculated conformation of the GC sequence is overall very similar to the crystallographic result. In the case of Z-RNA, the only crystallographic structure, obtained for the tetramer $BrGCBrGC$ (7), is compatible with the less stable conformations that we are located (large shift and intra strand hydrogen bond) calculated with 12 base pairs. The reason for this result may be in part due to hydration introduced by the bromine atoms. However, in order to test for oligomeric effects, we have also performed energy optimisations of the GC sequence using only 4 base pairs. For this fragment the stability of the large shift conformation becomes effectively more stable due to decreased inter-strand phosphate repulsions which decrease. In consequence, it should be stressed that structural studies of short oligomers should only be used with great caution in making deductions concerning longer stretches of nucleic acid, particularly when more than one conformational form can be adopted.

CONCLUSION

The molecular modelling carried out brings to light the fact that, within the Z- family, considerable polymorphism can exist. This polymorphism, while experimentally confirmed in solution studies has not yet been structurally characterized. In connection with our studies, it is particularly interesting to note that the chemical reactivity of the bases within natural Z-DNA sequences has been found to be highly variable and, for a particular base, may also evolve as a function of the superhelical density (6,17,18). By comparing our results with these variations, we have been able to suggest that the highly reactive zones within Z tracts may indeed be associated with large conformational discontinuities characterized by base pair shift changes , and,

furthermore, that the evolution of reactivity as a function of negative supercoiling can be understood if we take into account the fact that certain sequences can adopt more than one conformation with very little change of energy(12).

As concerns ribonucleic acid, our results suggest that the experimentally described Z_d and Z_r forms of Z-RNA (19,20) may well be explained in terms of the two stable conformations, again characterized by shift changes that we have found (13).

Comparisons with available experimental data thus appear to justify the existence of the polymorphism within the Z family indicated by our modeling study and also support a number of the detailed base sequence effects that we have described.

REFERENCES

- 1 E.B. Kmiec and W.K. Holloman, *Cell*, 44 (1986) 545-554.
- 2 R.A. Fischel, K. Detmer and A. Rich, *Proc. Natl. Acad. Sci. USA*, 85 (1988) 36-40.
- 3 A. Jaworski, W.T. Hsieh, J.A. Blaho, J.E. Larson and R.D. Wells, *Science*, 238 (1988) 773-777.
- 4 T.M. Jovin, L.P. McIntosh, D.J. Arndt-Jovin, D.A. Zarling, M. Robert-Nicoud and J.H. van der Sande in: B. Pullman and J. Jortner (Ed.), *Nucleic Acids : The Vectors of Life*, D.Reidel, Dordrecht, Holland, 1983, pp89-99. (S. Arnott and D.W.L. Hukins, *J. Mol. Biol.* 81 (1973) 93-105).
- 5 B. Hartmann, J. Ramstein and M. Leng, *FEBS Lett.*, 225 (1987).
- 6 N. Vogt, N. Rousseau, M. Leng and B. Malfoy, *J. Biol. Chem.*, 263 (1988) 11826-11832.
- 7 Y. Nakamura, S. Fujii, S. Urata, M. Itekara and K. Tomita, *Nucl. Acids Symp.*, Ser. n° 16 (1985) 29-32.
- 8 R. Lavery, H. Sklenar, K. Zakrzewska and B. Pullman, *J. Biomol. Struct. Dynam.*, 3 (1986) 989-1014.
- 9 R. Lavery and H. Sklenar, *J. Biomol. Struct. Dynam.*, 6 (1988) 63-91.
- 10 R. Lavery, *Structure and Expression Vol.3 DNA Bending and Curvature*, in: W.K. Olson, M.H. Sarma, R.H. Sarma and M. Sundaralingam (Eds.), *Adenine Press*, 1988, pp.191-211.
- 11 A.H.J. Wang, G.J. Quigley, F.J. Kolpak, J.H. van Boom, G.A. van der Marel and A. Rich, *Science*, 211 (1981) 171-176.
- 12 B. Hartmann, B. Malfoy and R. Lavery, *J. Mol. Biol.*, 207 (1989) 433-444.
- 15 B. Hartmann and R. Lavery, *J. Biomol. Struct. Dynam.*, in press
- 14 P.S. Ho, M.J. Ellison, G.J. Quigley and A. Rich, *EMBO J.*, 5 (1986) 2737-2744.
- 15 W. Saenger, *Principles of Nucleic Acid Structure*, Springer-Verlag, 1984, pp.69-78.
- 16 I.Jr. Tinoco, P.W. Davis, C.C. Hardin, J.D. Puglisi, G.T. Walker and J. Wyatt, *Cold Spring Harbour Symposia on Quantitative Biology vol LII*, 1987, pp.135-146.
- 17 W. Herr, *Proc. Natl. Acad. Sci. USA*, 82 (1985) 8009-8013.
- 18 B.H. Johnston and A. Rich, *Cell*, 42 (1985) 713-724.
- 19 M.O. Trulson, J.D. Puglisi, P. Cruz, I.Jr. Tinoco and Mathies, *Biochemistry*, 26 (1987) 8624-8630.
- 20 C.C. Hardin, D.A. Zarling, J.D. Puglisi, M.O. Trulson, P.W. Davies and I.Jr. Tinoco, *Biochemistry*, 26 (1987) 5191-5199.

MINIMIZATION AND MOLECULAR DYNAMICS OF Z-DNA MODIFIED BY ACETYLAMINOFLUORENE

V. FRITSCH and E. WESTHOF

Laboratoire de Cristallographie Biologique, Institut de Biologie Moléculaire et Cellulaire, Centre National de la Recherche Scientifique, 15 rue R. Descartes, 67084 Strasbourg Cedex, France.

SUMMARY

The modeling program AMBER 3.0 (ref. 1) was used to study the conformations adopted by the C8-substituted guanosine adduct of the carcinogen acetylaminofluorene (AAF), called dGuo-AAF. For that purpose, we have added to the program AMBER a subroutine allowing the use of a distance-dependent dielectric constant with the form suggested by Lavery *et al.* (ref. 2). One of the minimized conformations, compatible with the Z-DNA form, has been chosen and inserted into the hexamer $d(CGCGCG)_2$. The results of the minimizations and molecular dynamics (MD) simulations display a pronounced dependence on the choice of the dielectric constant and on the weight given to the 1-4 electrostatic contributions (scale factor SCCE in AMBER).

METHODS

The energy potential function used in our molecular mechanics and molecular dynamics simulations has the form

$$V(r) = V_b + V_\theta + V_\phi + V_{vdW} + V_{es} + V_{hb}$$

where r is the $3N$ dimensionnal vector specifying the Cartesian coordinates of the N atoms of the molecule. The first three terms correspond to deformations of the covalent structure, with contributions arising from the deformation of bonds (b), the deformation of bond angles (θ) and torsional rotations about bonds (ϕ). The last three terms correspond to the nonbonded interactions, broken into the van der Waals (vdW), the electrostatic (es) and the hydrogen bond (hb) contributions.

In this potential function, important errors arise from the electrostatic term, especially in the case of charged molecules like nucleic acids. To reduce such errors, explicit solvent molecules could be included in the calculations at the cost of much longer computation time and with possible convergence problems. If those molecules are not explicitly included, how to effectively simulate the solvent and charge screening effects ?

One approach uses a distance-dependent dielectric constant $\epsilon(r)$, where r is the interatomic distance (ref. 3, and references therein).

Another approach is the scaling of the partial atomic charges; the screening due to counterions in nucleic acids simulations is commonly represented by reducing the charges at phosphate groups (refs. 4-5).

But the difficulties of representing accurately the electrostatic effects in the polyanionic nucleic acids are great, and each of the previous molecular dynamics studies of DNA has adopted a different approach. The next table shows several of those approaches.

Approaches used in some MD simulations	References
electrostatic interactions are neglected; solvent molecules are not included.	(ref.4)
a distance-dependent dielectric constant, $\epsilon(r) = r$, is used; phosphate charges are scaled down to -0.2.	(ref.5)
simulations with fully charged phosphates, with counterions; distance-dependent dielectric constants, $\epsilon(r) = r$ and $\epsilon(r) = 4r$, are used.	(ref.6)
simulations with counterions; solvent molecules are included explicitly; a dielectric constant, $\epsilon = 1$, is used.	(ref.7)
solvent molecules are included explicitly; a dielectric constant, $\epsilon = 1$, is used; scaling of atomic partial charges so that the system is electrically neutral.	(ref.8)

In the table, the distance-dependent dielectric constants are $\epsilon(r) = r$ and $\epsilon(r) = 4r$. But others functions have been proposed. We have studied two of them:

$$\epsilon(r) = A + \frac{B}{1 + ke^{-\lambda Br}}$$

where $A = -20.929$, $B = 99.329$, $\lambda = 0.001787$, $k = 3.4781$ (ref. 9).

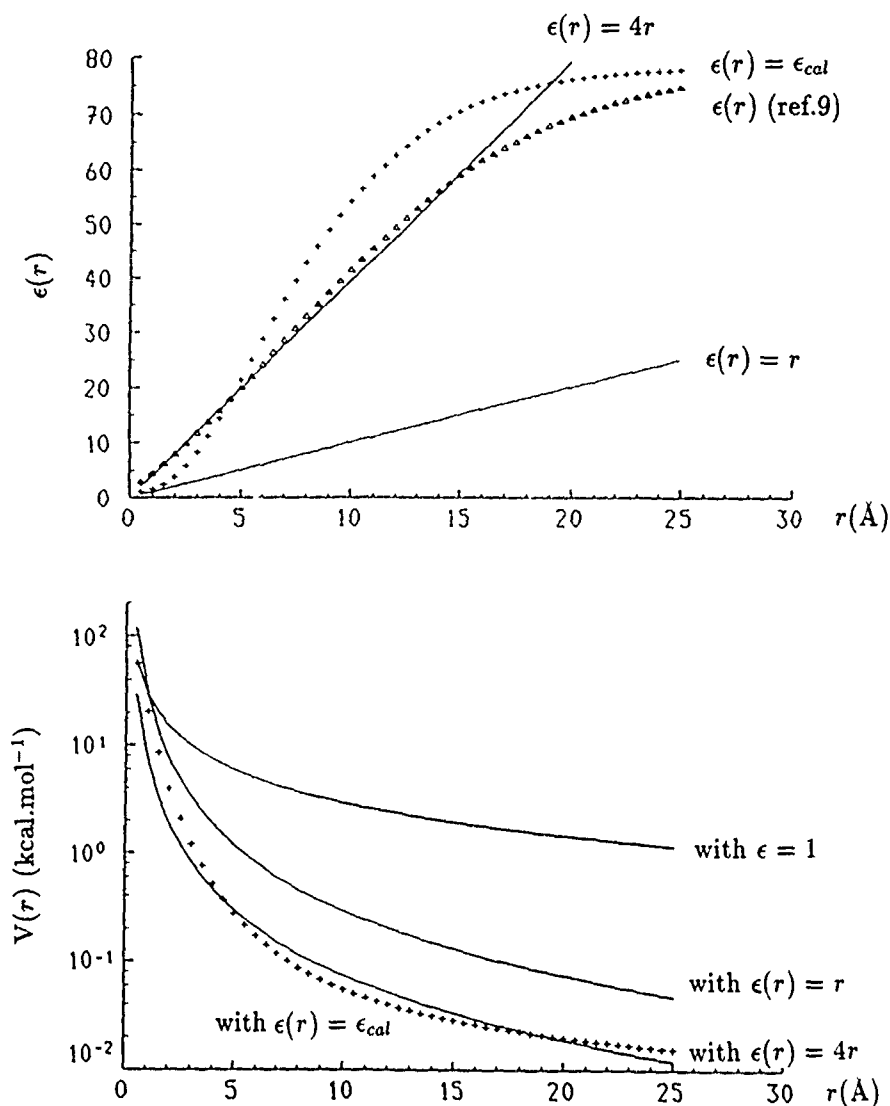
$$\epsilon(r) = D - \frac{D-1}{2}((Ar)^2 + 2Ar + 2)e^{-Ar}$$

where $D = 78$, $C = 2.674$, $H = 7.5$, $A = \frac{C}{H}$ (ref. 2)

The diagrams below represent the variations of different distance-dependent dielectric constants $\epsilon(r)$ and the variations of the associated electrostatic potentials (in kcal.mol⁻¹) as a function of the interatomic distance r

$$V(r) = \frac{332q_i q_j}{\epsilon(r)r}$$

with $q_i = q_j = 0.3$ units of proton charge .

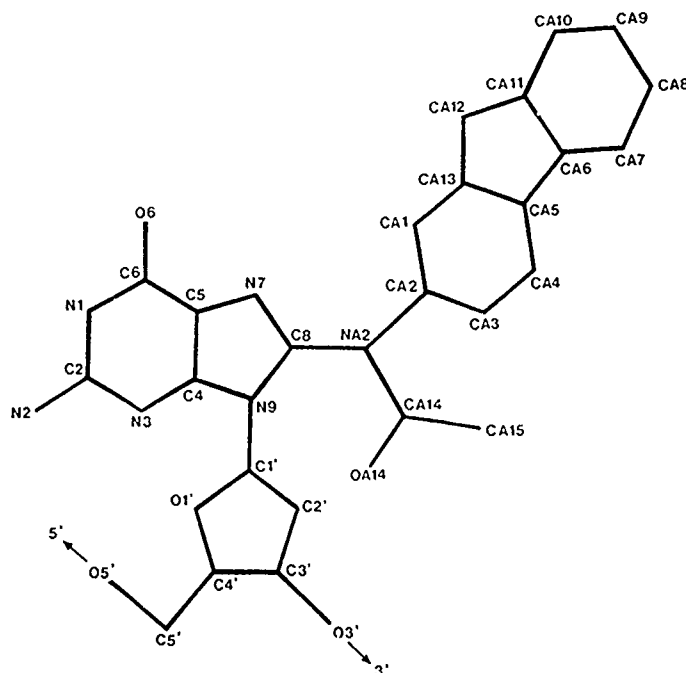


Those figures show that $\epsilon(r) = 4r$, for $r < 20 \text{ \AA}$, is a good approximation of the sigmoidal functions. On the other hand, electrostatic effects appear over-estimated using $\epsilon(r) = r$, especially at distances larger than the Debye length (8-10 \AA) and in the case of helical nucleic acids where most atoms are accessible to solvent molecules.

The function developed by Lavery *et al.* (ref. 2), called ϵ_{cal} , has the advantage of being more sensitive for small values of interatomic distance r , which increases the accuracy of the calculations. Moreover, first and second derivatives are continuous, essential conditions for energy-minimization algorithms. Therefore, we decided to use the latter dielectric function in the electrostatic energy calculations. For this purpose, the program AMBER was modified and a subroutine was added allowing the use of this distance-dependent dielectric constant.

RESULTS

It has been proposed that conformational properties of the AAF-bound region in a double-stranded DNA molecule are important factors in the process of carcinogenesis (ref. 10). We first studied the conformational and dynamics properties of the C8 bound guanosine adduct of AAF (dGuo-AAF). The molecular mechanical parameters reported by Weiner *et al.* (ref. 1) were used. Additional parameters and charges were necessary for AAF (to be published). Coordinates of the guanine AAF adduct are from Neidle *et al.* (ref. 11). The nomenclature used for dGuo-AAF is shown below.

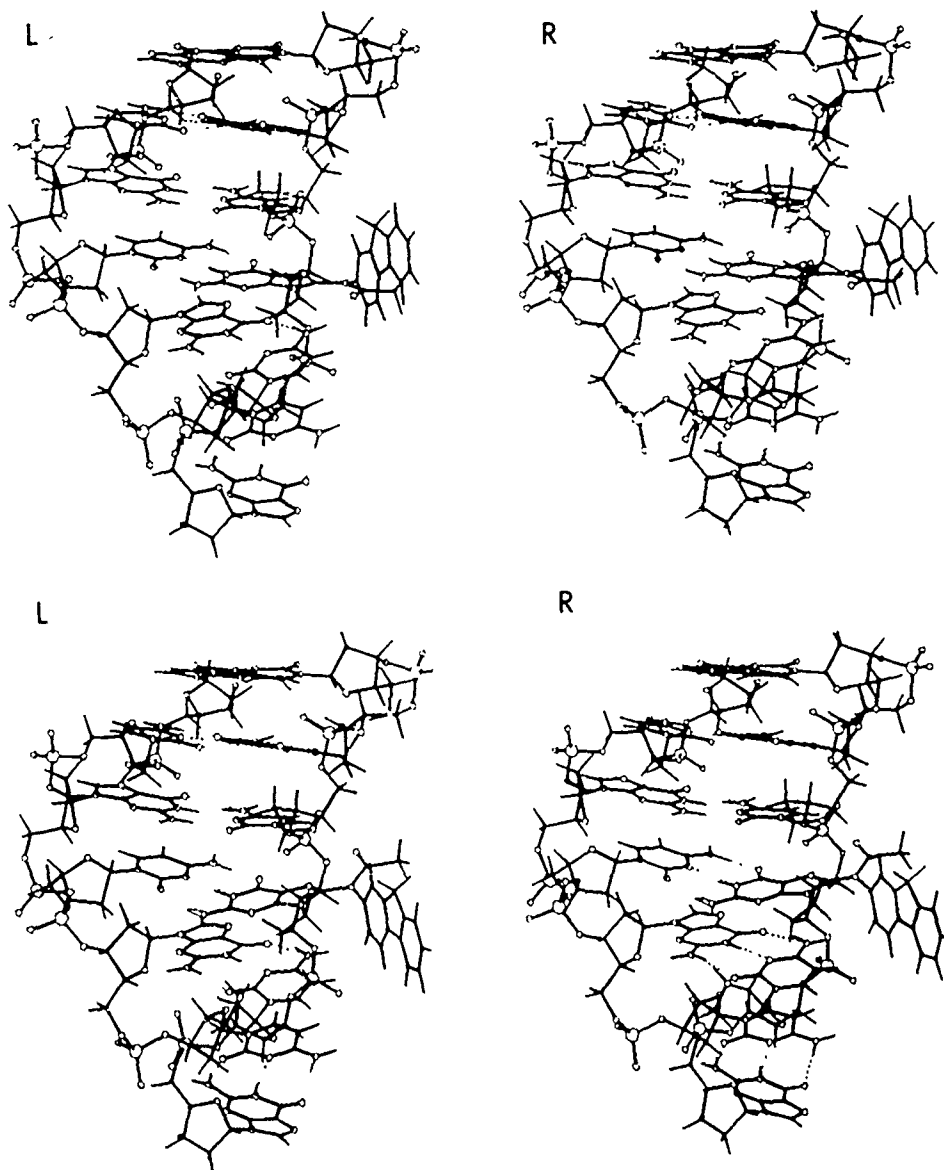


For the conformational study, systematic minimizations were done in dependence of two dihedral angles, defined by N9-C8-NA2-CA2 for α and O1'-C1'-N9-C8 for χ , which varied between 0° and 360° in step of 30° . The angle γ , defined by C8-NA2-CA14-CA15, was kept near 180° in agreement with the results of Evans and Miller (ref. 12).

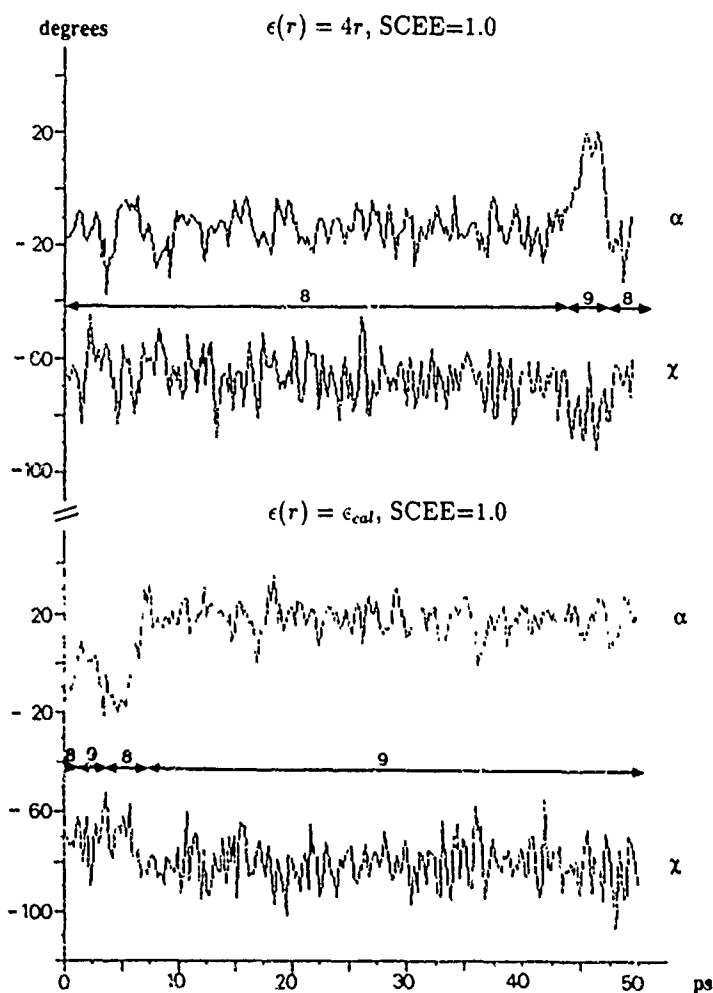
Calculations (with ϵ_{cal} , SCEE=1.0) gave us a minimization map with twelve main minima. Molecular dynamics runs (300 K, 50 ps), starting from each minimum, revealed that the four domains centered on $\alpha = 0^\circ$ and $\chi = -100^\circ$ were connected by low conformational barriers. During molecular dynamics in each of the twelve domains, the dGuo-AAF sugar ring adopted preferentially the C2'-endo pucker but the pseudorotation phase angle extended throughout the southern hemisphere of the pseudorotation circle (from O1'-endo to O1'-exo). The amplitude of pucker, however, remained centered on 39° .

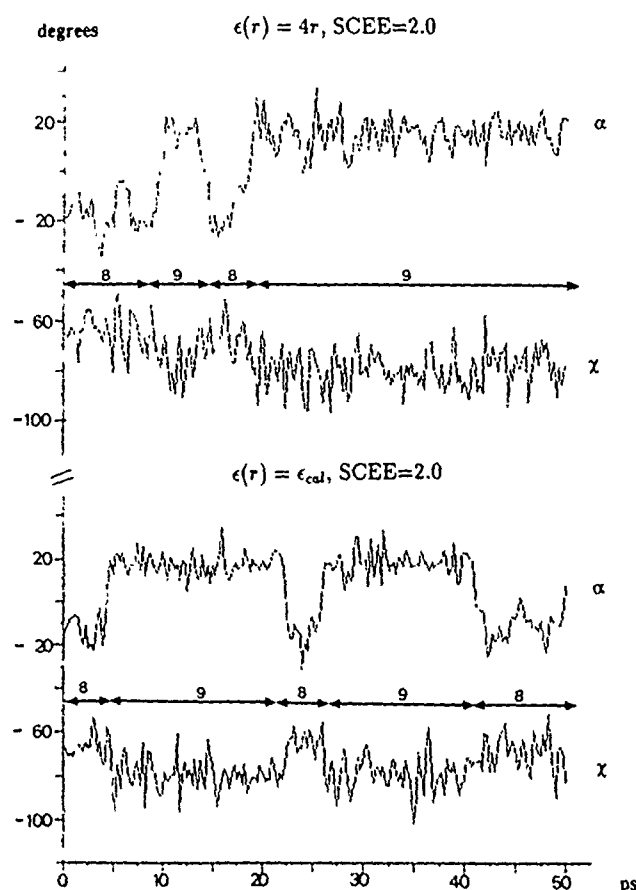
With the program FRODO (ref. 13), the conformation of guanosine G4 in the hexamer $d(CGCGCG)_2$ (coordinates are from Wang *et al.* (ref. 14)) was replaced by each of the two minimized conformations of dGuo-AAF compatible with Z-DNA structure.

Stereo views below show examples of the conformations of the AAF-modified hexamer after minimization (with ϵ_{cal} , SCEE=1.0). For the first stereo view, corresponding to domain 9, values of angles α and χ are 18° and -87° . For the second, corresponding to domain 8, the values are -13° and -64° .



An important result was the formation, after minimization in domain 9 (but only with ϵ_{cal}), of a hydrogen bond between the oxygen OA14 of AAF and a hydrogen bound to the nitrogen N4 of the cytosine C5. Minimizations starting from domain 8 (whatever the electrostatic parameters) led to a conformation where such an interaction is impossible. Such an intramolecular H-bond between bound AAF and Z-DNA has never been noticed. In order to study its stability, molecular dynamics simulations were performed starting from minimizations of the hexamer in domain 8 with only residues C3-G4-C5 shaken during the dynamics by the use of the belly option. The simulations were done with different distance-dependent dielectric constants for two values of the scale factor for 1-4 electrostatic interactions (SCEE). Here, we present the results obtained by using $\epsilon(r) = \epsilon_{cal}$ and $\epsilon(r) = 4r$, with SCEE equal to 1 or 2. As shown in the diagrams below, which display the variation of torsional angles α and χ between domain 8 (no H-bond) and domain 9 (with H-bond) during MD runs, the stability of the intramolecular H-bond between OA14(AAF) and H-N4(C5) depends strongly on the choice of the electrostatic parameters.





Thus, with $\epsilon(r) = 4r$ and $\text{SCEE}=1.0$, the AAF remained most of the simulation in domain 8; while with $\epsilon(r) = \epsilon_{cal}$ and $\text{SCEE}=1.0$, the AAF flipped soon to domain 9 and remained there with formation of the intramolecular H-bond. With $\text{SCEE}=2.0$, the oscillations between the two domains are more frequent and the conformation with the intramolecular H-bond seems favoured. However, independently of the parameters chosen during the MD simulations, the dGuo-AAF sugar pucker adopted the C_4' -endo- $O1'$ -exo conformation with an amplitude of 39° . In contrast, during MD simulations of the unmodified Z-hexamer, large variations in phase and amplitude of pseudorotation of the guanine sugars were observed (ref. 15).

In conclusion, for charged molecules like nucleic acids, electrostatic parameters influence greatly the preferred conformations and their dynamics. Minimizations and MD simulations should be run systematically with different values for those parameters.

It is known that covalent binding of AAF to poly d(G-C) stabilizes the Z conformation (ref. 16). The intramolecular H-bond between AAF covalently linked to guanosine C8 and the amino group of the 3'-proximal cytosine revealed by the present calculations contributes probably to the stability of AAF-modified Z-DNA.

REFERENCES

- 1 S.J. Weiner, P.A. Kollman, D.T. Nguyen and D.A. Case, An all atom force field for simulations of proteins and nucleic acids, *J. of Comp. Chem.*, 7 (1986) 230-252
- 2 R. Lavery, H. Sklenar, K. Zakrzewska and B. Pullman, The flexibility of the nucleic acids: (II) the calculation of internal energy and applications to mononucleotide repeat DNA, *J. Biomol. Struct. Dyn*, 3 (1986) 989-1014
- 3 N.K. Rogers, The modelling of electrostatic interactions in the function of globular proteins, *Prog. Biophys. molec. Biol.*, 48 (1986) 37-66
- 4 M. Levitt, Computer simulation of DNA double-helix dynamics, *Cold Spring Harbor Symp. Quant. Biol.*, 47 (1983) 251-262
- 5 B. Tidor, K.K. Irikura, B.R. Brooks and M. Karplus, Dynamics of DNA oligomers, *J. Biomol. Struct. Dyn.*, 1 (1983) 231-252
- 6 P. Kollman, J.W. Keepers and P. Weiner, Molecular - mechanics studies on $d(CGCGAATTCGCG)_2$ and $dA_{12}.dT_{12}$: an illustration of the coupling between sugar repuckering and DNA twisting, *Biopolymers*, 21 (1982) 2345-2376
- 7 G.L. Seibel, U.C. Singh and P.A. Kollman, A molecular dynamics simulation of double-helical B-DNA including counterions and water, *Proc. Natl. Acad. Sci. USA*, 82 (1985) 6537-6540
- 8 P.S. Subramanian and D.L. Beveridge, A theoretical study of the aqueous hydration of canonical B $d(CGCGAATTCGCG)$: Monte Carlo simulation and comparison with crystallographic ordered water sites, *J. Biomol. Struct. Dyn*, 6 (1989) 1093-1122
- 9 E.L. Mehler and G. Eichele, Electrostatic effects in water-accessible regions of proteins, *Biochemistry*, 23 (1984) 3887-3891
- 10 J.H. Nelson, D. Grunberger, C.R. Cantor and I.B. Weinstein, Modification of ribonucleic acid by chemical carcinogens. IV. Circular dichroism and Proton Magnetic Resonance studies of oligonucleotides modified with N-2- acetylaminofluorene, 62 (1971) 331-346
- 11 S. Neidle, R. Kuroda, S. Broyde, B.E. Hingerty, R.A. Levine, D.W. Miller and F.E. Evans, Studies on the conformation and dynamics of the C8-substituted guanine adduct of the carcinogen acetylaminofluorene; model for a possible Z-DNA modified structure, *Nucleic Acids Res.*, 12 (1984) 8219-8233
- 12 F.E. Evans and D.W. Miller, Conformation and dynamics of carcinogenic N- substituted 2-aminofluorene compounds studied by Nuclear Magnetic Resonance spectroscopy, *J. Am. Chem. Soc.*, 105 (1983) 4863-4868
- 13 T.A. Jones, A graphic model building and refinement system for macromolecules, *J. Appl. Cryst.*, 11 (1978) 268-278
- 14 A.H.-J. Wang, G.J. Quigley, F.J. Kolpak, J.L. Crawford, J.H. van Boom, G. van der Marel and A. Rich, Left-handed double helical DNA: variations in the backbone conformation, *Science*, 211 (1981) 171-176
- 15 E. Westhof, B. Chevrier, S.L. Gallion, P.K. Weiner and R.M. Levy, Temperature-dependant molecular dynamics and restrained X-Ray refinement simulations of Z-DNA hexamer, *J. Mol. Biol.*, 190 (1986) 699-712
- 16 E. Sage and M. Leng, Conformation of poly(dG-dC).poly(dG-dC) modified by the carcinogens N-acetoxy-N-acetyl-2-aminofluorene and N-hydroxy-N-2-aminofluorene, *Proc. Natl. Acad. Sci. USA*, 77 (1980) 4597-4601

**A THEORETICAL INVESTIGATION OF THE BASE SEQUENCE PREFERENCES OF
MONOINTERCALATING POLYMETHYLENE CARBOXAMIDE DERIVATIVES OF
9-AMINOACRIDINE.**

Christian COULOMBEAU* and Nohad GRESH**

*Université J. Fourier, LEDSS VI, B.P. 53F, 38041 Grenoble,
France, and

**Laboratoire de Biochimie Théorique associé au C.N.R.S.,
Institut de Biologie Physico-Chimique, 13, rue Pierre et
Marie Curie, 75005 Paris, France.

ABSTRACT

Theoretical computations are performed of the comparative binding affinities of five polymethylene carboxamide derivatives of 9-aminoacridine to a series of double-stranded hexanucleotides. The purpose of this investigation is to ascertain whether minor groove recognition of a guanine base adjacent to the intercalation site can occur, and be preferentially stabilized, for a given length of the polymethylene side chain, encompassing from $n=2$ up to $n=6$ methylene groups. For that purpose, several representative sequences were investigated, in which intercalation of the 9-aminoacridine chromophore occurred at a central d(CpG) or d(TpA) step. Investigated were the self-complementary sequences d(GCGGCG)₂, d(GCCGGC)₂, d(TATATA)₂ and d(ATAAT)₂, as well as the "mixed" sequences d(ACTAAT).d(ATTAGT) and d(TGTATA).d(TATACA). For $n=3$ up to $n=6$, such a recognition was enabled only when the guanine base was located downstream of the intercalation site, i.e. with steps d(CGG) and d(TAG). It occurred by means of a bidentate interaction involving, on the one hand, H(N₂) and N₃ of the base, and, on the other hand, the carbonyl oxygen and the *cis* amino hydrogen of the terminal formamide moiety of the ligand. Because of the flexibility of the side chain, however, alternative binding modes were also found to occur competitively, involving backbone-only interactions of the side chain.

On the basis of the present computations using the SIBFA procedure [1-2], upon binding to the sequence d(GCCGGC)₂, an optimal value of $n=5$ could be derived, with the corresponding acridine derivative eliciting both a significant prevalence of the bidentate over backbone only binding mode, and the most favourable energy balance within the investigated series (see Table I). This privileged value of $n=5$ is fully consistent with the experimental results of Markovits et al. and Gaugain et al. [3-5]. The very flexibility of the side chain, however, hampered any preferential recognition of a triplet sequence with a downstream guanine, such as d(CGG) or d(TAG), to be elicited over sequences such as d(TAA), d(TAT) or d(TAC).

A more complete account has been submitted for publication by the authors to *Nucleic Acids Research*.

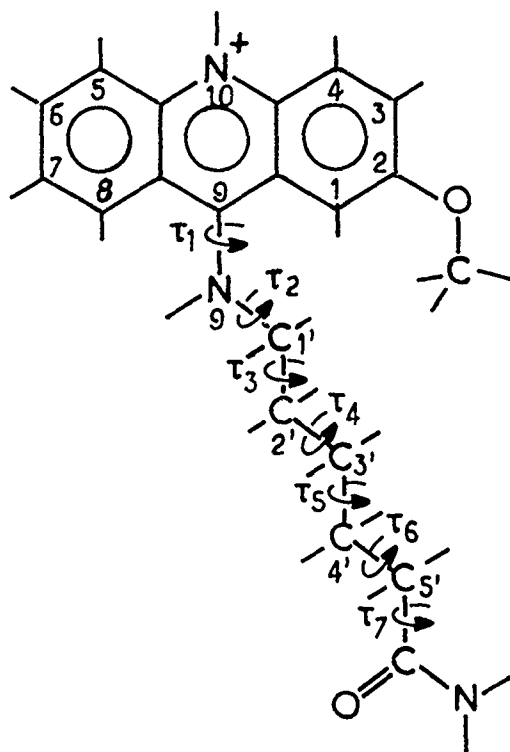


Fig. 1 The investigated polymethylene carboxamide derivatives of 2-methoxy-9-aminoacridine. Structural formula of the representative compound with five methylene groups.

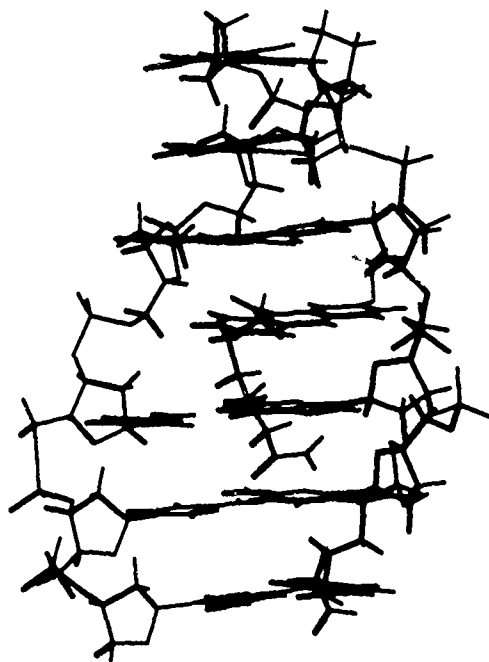


Fig. 2 Representation of the bidentate complex of compound N5 with $d(GCCGGC)_2$.

TABLE I

Values of the binding energetics in the energy-minimized complexes of N2-N6 with the self-complementary sequence d(GCCGGC)₂.
Energies in kcal/mole.

d(GCCGGC) ₂ - Bidentate mode				
	N3	N4	N5	N6
ΔE	-252.4	-254.4	-260.0	-256.6
δlig	19.6	12.3	14.8	18.9
δE	-232.8	-242.1	-245.2	-237.7
δ	12.4	3.1	0.0	7.5

d(GCCGGC) ₂ - Nonbidentate binding mode					
	N2	N3	N4	N5	N6
ΔE	-243.6	-240.0	-251.3	-258.0	-258.5
δlig	9.9	17.7	16.0	16.3	17.4
δE	-233.7	-222.3	-235.3	-241.7	-241.1
δ	11.5	22.9	9.0	3.5	4.1

REFERENCES

- (1) Gresh, N., Claverie, P. and Pullman A. (1984) Theoret. Chim. Acta 66, 1-20.
- (2) Gresh, N., Claverie, P. and Pullman, A. (1986) Int. J. Quantum Chem. 29, 101-118
- (3) Markovits, J., Gaugain, B., Roques, B.P. and Le Pecq, J.B. (1981) Jerus. Symp. Quantum Chem. Biochem. 14, 285-298.
- (4) Gaugain, B., Markovits, J., Le Pecq, J.B. and Roques, B.P. (1981) Biochemistry, 20, 3035-3052.
- (5) Markovits, J., Gaugain, B., Barbet, J., Roques, B.P. and Le Pecq, J.B. (1981) Biochemistry 20, 3042-3048.

**NMR STUDY AND MOLECULAR MODELLING
OF TWO DOUBLE STRANDED DNA FRAGMENTS
TGACGTCA AND ACTGCAGT**

J.M. PIRIOU¹, O. MAUFFRET¹, J. ARMIER², S. FERMANDJIAN¹, M. LE BRET¹

¹ Laboratoire de Biochimie-Enzymologie CNRS URA 158, INSERM U140
Institut Gustave-Roussy- 39, rue Camille Desmoulins
94805 Villejuif Cedex FRANCE

² Biologie et Génétique Moléculaires CNRS UPR 2
7, rue G. Moquet - 94802 Villejuif Cédex FRANCE

SUMMARY

Footprinting experiments showed that the antitumour drug N-methyl 9-hydroxy ellipticine (NMHE) preferentially binds to CG sites in double stranded DNA, and led to the determination of a consensus sequence ACGT (B. RENE and co-workers).

To elucidate this selectivity, we have carried out the ¹H NMR study and the molecular modelling of the binding of NMHE to a DNA octamer containing the consensus sequence -TGACGTCA -, and its inverse sequence ACTGCAGT. The two fragments, which are both self-complementary, have first been studied without the drug.

From the coupling constants J1'2', J1'2'', J2'3', J2''3', measured from COSY experiments, we have considered each sugar ring as the result of an equilibrium between the C2'-endo and the C3'-endo conformations, which can be represented by the percentage of C2'-endo form (1). We also carried out constrained minimizations, using the experimental couplings and interproton distances (H6/H8 - H1', H2', H2'', inter- and intra-nucleotide), determined from NOESY experiments.

The work concerning the binding of the drug to the fragments is currently being done.

MATERIALS AND METHODS

Oligonucleotides TGACGTCA (I) and ACTGCAGT (II) have been synthesized at the I.R.S.C. (Villejuif) using the solid phase procedure on an Applied Biosystems 380 B automatic apparatus. Final purification has been ensured by HPLC and purity checked by NMR. Samples have been prepared in deuterated phosphate buffers (pH 7, I = 0.1 and 0.2 mM EDTA. Oligonucleotides were lyophilized twice in ²H₂O.

Nuclear Magnetic Resonance : data collection and processing.

Two dimensional proton NMR spectra were collected on Bruker spectrometers at various fields : 500 MHz (AM500 ; Laboratoire de Chimie Organique, Jussieu), and 300 MHz (MSL300 ; I.R.S.C. Villejuif).

Coupling constants : ¹H-¹H vicinal couplings constants J1'2' and J1'2'' were measured on well-resolved H1' resonances in 1D spectra ; J2'3' and J2''3' were extracted from COSY phases with double quantum filter recorded with high resolution in F₂ dimension (12 Hz/pt), as mentioned elsewhere (1).

Distance constraints. Two dimensional phase sensitive NOESY spectra in $^2\text{H}_2\text{O}$ were recorded at 300 MHz for mixing times of 100, 150 and 200 ms. The data in the t_1 and t_2 dimensions were apodized by using a 90° phase-shifted sine bell before Fourier transformation. Volumes for each resolved cross-peak were measured from 2D NOESY spectra. A cytidine H5-H6 distance of 2.45 Å was used as a yardstick for all NOEs between nonexchangeable protons.

Molecular mechanics calculations and graphics manipulations were carried out on a Silicon Graphics IRIS 4D 70/GT workstation, using the program AMBER (2) (Kollman *et al.*) associated to an energy minimizer, ORAL, developed in the laboratory, and the program NACAD (Nucleic Acid Computer-Aided Design), developed in the laboratory as well.

The parameters were those described by Weiner *et al.* (3) (1986) and by Singh *et al.* (4) (1986). All hydrogen atoms were treated explicitly. To simulate the screening effect of the solvent a gas phase potential was employed, where the dielectric constant D_{ij} is proportional to the distance d_{ij} separating a pair of atoms : $D_{ij} = C \cdot d_{ij}$ (Gelin & Karplus, 1981 ; Weiner *et al.*, 1984) (5,6). C was taken as 4 Å^{-1} . All pairs were included in the calculation of nonbonding interactions. The minimizations were carried out with the 1-4 interatomic interactions divided by two. Refinements were terminated when the norm of the energy gradient was less than 0.01 kcal/Å.

RESULTS AND DISCUSSION

From the observation of the experimental coupling constants, given in table 1., one can see that the two fragments present significant differences in the geometries (puckering) of the sugars.

(I)	$J_{1'2'}^a$	$J_{1'2''}^a$	$J_{2'}^b$	$J_{2''}^c$
T1	8.6	5.2	29.0	20.2
G2	10.1	5.5	26.6	21.1
A3	9.3	5.3	28.0	21.1
C4	8.5	5.8	30.5	23.7
G5	8.2	6.2	28.0	22.7
T6	7.3	7.3	30.5	22.7
C7	7.0	7.0	29.0	24.0
A8	6.8	6.8	28.0	24.2

(II)	$J_{1'2'}^a$	$J_{1'2''}^a$	$J_{2'}^b$	$J_{2''}^c$
A1	7.8	6.2	28.0	20.8
C2	7.3	7.3	28.0	22.0
T3	8.5	6.1	29.0	23.0
G4	9.6	5.8	26.5	23.0
C5	7.8	6.1	28.0	22.0
A6	8.7	6.2	30.0	22.0
G7	7.3	7.3	28.0	22.5
T8	7.0	7.0	^d	^d

Table 1. ^1H - ^1H coupling constants for octamers (I) and (II).

^a : measured on 600 MHz 1D spectra at 30 °C except for T1 and C7 measured at 24 °C (accuracy ± 0.365 Hz) and C2 and G7 which were measured on the H1'/H2' COSY cross-peak (accuracy ± 1.2 Hz). ^b : measured on the H1'/H2' COSY cross-peak (accuracy ± 1.2 Hz). ^c : measured on the H1'/H2'' COSY cross-peak (accuracy ± 1.2 Hz). ^d : H2' and H2'' strongly overlap.

For both molecules a set of $64 = 2^6$ starting structures was generated, considering for each sugar ring (except those at the extremities) the C2'-endo and C3'-endo conformations. Those models were refined, without constraining the geometry of the deoxyribose. We selected the best ten ones in terms of energy (with less than 5 kcal difference between one another). We calculated the coupling constants of our models and confronted them to the experimental constants. We considered each nucleotide independently. For most of them, the calculated couplings were in good agreement with the experimental ones (pure C2'-endo). For the other ones - A₃, C₄, G₅ -, we had to consider the equilibrium C2'-endo - C3'-endo, i.e. taking the C2'-endo form of one sugar, we tried to find another structure where the same sugar exists in the C3'-endo conformation, and a fraction p such as, for the 4 couplings :

$$p.J_{C2'endo} + (1-p).J_{C3'endo} = J_{exp} \quad (1)$$

We ruled out the structures for which it was not possible to find such a fit for all nucleotides.

It came out that no model was left at the end. Therefore constrained minimizations were run in order to achieve the determination of the structures with another method. Experimental

Model	E (kcal/mol)	$\sum (d_{calc} - d_{exp})^2 / d_{calc}$	$\sum (J_{calc} - J_{exp})^2 / J_{calc}$
Arnott (minimized)	-196	3.7	18.7
Arnott with constraints on sugar torsion angles	-192	3.3	15.6
Arnott with constraints on interproton distances	-200	3.2	18.6
(2) + (3)	-198	3.4	21.5
Modified twist+pucker from Rinkel & Altona without constraints	-206	3.7	13.9
(5) with constraints on the delta angle	-189	3.3	16.0

Model	E (kcal/mol)	$\sum (J_{calc} - J_{exp})^2 / J_{calc}$
Arnott (minimized)	-188	16.7
Arnott with constraints on sugar torsion angles	-188	13.5
Modified twist+pucker from Rinkel & Altona without constraints	-192	12.6
(3) + constraints on the delta angle	-192	13.1

Table 2. a : Energies and sums of errors on the 54 distances and 24 couplings (J1'2', J1'2'', J2'3' and J2'3'') for several refined models (TGACGTCA).

b : Energies and sums of errors on the 24 couplings J1'2', J1'2'', J2'3' and J2'3'' - (ACTGCAGT).

a	Step	Theoretical value (Sarai <i>et al.</i>)	Helical Twist for the model
	T-G	29	37.5
	G-A	38	34.5
	A-C	40	39.7
	C-G	23	35.7
	G-T	40	39.7
	T-C	38	34.5
	C-A	29	37.5

b	Step	Theoretical value (Sarai <i>et al.</i>)	Helical Twist for the model
	A-C	40	40.9
	C-T	35	34.3
	T-G	29	33.8
	G-C	38	40.3
	C-A	29	33.8
	A-G	35	34.3
	G-T	40	40.9

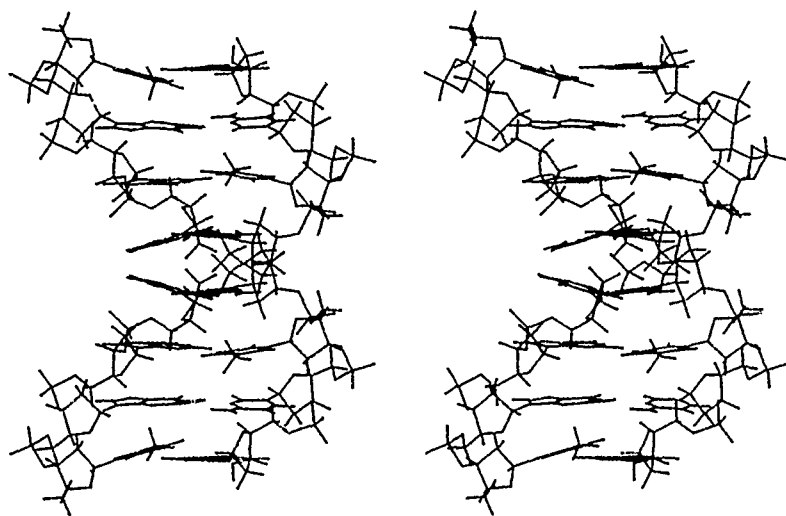
Table 3. Helical Twist for the refined models obtained from Arnott structures with modified twists : a- TGACGTCA, b- ACTGCAGT.

couplings (24 constants for each fragment) were converted into torsion angles, using the generalized Karplus equation (7), and introduced as constraints in the calculations, as well as 54 interproton distances (for TGACGTCA only). Different combinations were used, as shown in tables 2. : no constraints, constraints on the torsion angles and/or on the distances, with different starting structures.

The most satisfactory model came from the calculations using constraints, and, interestingly, as shown in table 2., modified helical twist. Indeed we applied to an Arnott structure (8) where the puckers had already been set equal to the values found thanks to the method exposed by Rinkel & Altona, the theoretical helical twists determined by Sarai *et al.* (9). Then we refined those starting structures with and without constraining the delta angle which determines the sugar puckering. Stereoscopic views of the two fragments (refined structures with the modified twist and without constraining delta) are given in *fig. 1*. and the helical twists are shown in table 3. It is also interesting to note that in all calculations the symmetry of the two strands of the fragments is conserved. More over one can see in table 3. that in both fragments the twist is "periodic", with a period of 2 in the case of TGACGTCA and a period of 3 for ACTGCAGT. Nevertheless the values of the twist calculated for the refined models differ significantly from those described by Sarai *et al.* (9).

The results proved to be energetically along the most favourable and in good agreement with the experimental couplings and interproton distances.

a



b

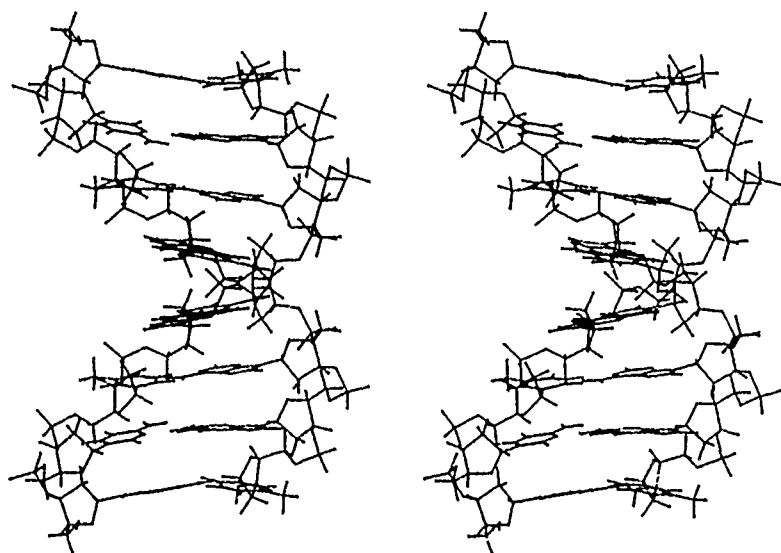


Fig. 1. : Stereoscopic views of TGACGTCA (a) and ACTGCAGT (b). The models represented here correspond to the refined structures with modified twists and puckers.

CONCLUSION

In agreement with NMR results, molecular mechanics calculations have shown that the two octameric oligonucleotides present significantly different structures :

- in terms of sugar puckering : whereas TGACGTCA presents a certain variety of puckers in the range C2'-*endo*, ACTGCAGT appears more homogenous ($P = 150^\circ$ for all nucleotides) ;
- in terms of helical parameters : ACTGCAGT presents larger variations in the helical twist than TGACGTCA.

We are now investigating other parameters (propeller twist...) in order to get further details on the fine structure of the two octamers, which may help understand their recognition by the drug NMHE.

REFERENCES

- 1- RINKEL J.L., ALTONA C. ; Conformational analysis of the deoxyribofuranose ring in DNA by means of sums of proton-proton coupling constants : a graphical method. *J. Biomol. Struct. and Dynam.* (1987) 4 621-649
- 2- WEINER P.K., KOLLMAN P.A. ; AMBER ; Assisted Model Building with Energy Refinement. A General Program for Modelling Molecules and their interactions *J. Comp. Chem.* (1981) 2 287-303
- 3- WEINER S.J., KOLLMAN P.A., NGUYEN D.T., CASE D.A. ; An All Atom Force Field for Simulation of Proteins and Nucleic Acids. *J. Comp. Chem.* (1986) 7 230-252
- 4- WEINER S.J., KOLLMAN P.A., CASE D.A., CHANDRA SINGH U., GHIO C., ALAGONA G., PROFETA Jr S., WEINER P. ; A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins *Jal. Am. Chem. Soc.* (1984) 106 765-784
- 5- GELIN B., KARPLUS M. *Proc. Nat. Acad. Sci. USA.* (1981) 77 801-805
- 6- SINGH U.C., WEINER P.K., CALDWELL J.W., KOLLMAN P.K. ; AMBER 3.0 University of California, San Francisco 1984
- 7- HAASNOOT C.A.G., DE LEEUW F., ALTONA C. ; The relationship between proton-proton NMR coupling constants and substituent electronegativities ; an empirical generalization of the Karplus equation. *Tetrahedron* (1980) 36 2783-2792
- 8- ARNOTT S., CAMPBELL-SMITH P., CHADRESEKHARAN P. ; CRC Handbook of Biochemistry (1976) 2 411-441
- 9- SARAI A., MAZUR J., NUSSINOV R., JERNIGAN R.L. ; Origin of DNA helical structure and its sequence dependence. *Biochem.* (1988) 27 8498-8502.

NUCLEOTIDE MISPAIRS STABILIZED BY WATER BRIDGES.

MODELING OF STRUCTURE AND ROLE IN TEMPLATE BIOSYNTHESIS.

V.I. Poltev, S.V. Steinberg

Institute of Biological Physics USSR Academy of Sciences
Pushchino, Moscow Region, 142292, USSR

SUMMARY

To explain the experimental data on the structure of double helical oligonucleotides with mispairs, to study the molecular mechanisms of point errors in the template directed nucleic acid biosynthesis, to create the conformation theory of wobble at the codon-anticodon recognition, we have examined the possibility of formation of mispairs via water bridges. Such a water molecule forms H-bonds with both bases of the same pair. For every combination of nucleotides several energy minima corresponding to two bases being in the same plane and linked with the same water molecule were found. Other water molecules can additionally stabilize some of these pairs. Theoretical conformational analysis of DNA fragments containing some of these pairs has permitted to propose this mechanism to play a role in arising of errors under nucleic acids biosynthesis. Calculations of the system "codon + anticodon loop + water molecule" reveal energy minima for every permissible wobble mispair in the third position of codon - anticodon complex. In these calculations codon was in A-RNA conformation, and the configuration of anticodon loop was rather similar to that of yeast Phe-tRNA.

INTRODUCTION

Besides Watson-Crick A:T(A:U) and G:C pairs nucleic acid bases can form many other "incorrect" planar pairs. Due to formation of such mispairs many point errors in replication, repair, transcription, and translation are observed. Moreover such pairs appear in the third position of codon-anticodon complex and in the tertiary structure of RNA. Thus the investigation of all the possibilities of formation of planar base pair is very important. There exist minima of energy of intermolecular interactions for every couple of nucleic acid bases being in usual tautomeric forms that allow to fit such a pair within the double helix (ref. 1-3). Appearance of errors both spontaneous and base analog induced during the nucleic acid biosynthesis in many cases can be explained by the formation of these pairs. However for some base combinations the pairs which have been constructed by means of water bridges are apparently more likely than pairs without such bridges. The mutual arrangement of the bases in some of such pairs (particularly in pyrimidine-pyrimidine pairs) is closer to that in Watson-Crick pairs than in pairs without water molecules.

This paper is devoted to the search of energy minima of the systems consisting of two bases and one or two water molecules and to calculations of low energy conformations of double helical DNA and RNA fragments containing some pairs with water bridges. The results of calculations are applied to account for experimental data concerning structure of oligonucleotide duplexes with mispairs in solution, to elucidate the mechanisms of point error arising during nucleic acid biosynthesis and to create the conformational theory of wobble upon the codon-anticodon recognition. Some of the results presented here were published in more detail earlier (ref. 4-7).

METHODS

Potential functions.

The calculations of the energy of intra- and intermolecular interactions were performed by the method of atom-atom potential functions using the procedures and parameters described and used earlier (ref. 2, 5-8). The total energy was calculated as a sum of the energies of all atom-atom interactions. The dependencies of the energy of interactions on the distances between the atoms were approximated by (1-10-12)-potentials for interactions of hydrogen atoms bound to N or O and proton-acceptor atoms (N or O) or by (1-6-12)-potentials for all other interactions. The phosphate groups were neutralized, i.e. the charges of each of two phosphate oxygens were diminished by a half of electron charge. Torsional potentials for rotation around single bonds and the energy of sugar ring bond angle distortion are also taken into account in calculation of intramolecular nonbonded interactions.

The calculated systems and the variable parameters.

(i) The base pairs. The systems contained two bases and one (or two) water molecule. Energy of the system was calculated as a function of 11 (or 17) variables where 5 variables determined the position of one base relatively to the other and 6 (or 12) - the position of one (or two) water molecules. As a rule both bases and water oxygens were in the same plane. The increasing of the number of variables (including propeller twist and buckle of the bases, permission of displacement of water oxygens out from the plane) did not result in considerable displacement of position of the energy minimum. The initial mutual positions of the bases and water molecules were determined by means of space filling models and after that the energy minimization was performed. When energy minimum had been achieved, the distance R between C1' pentose atoms of nucleotide pair of imaginary duplex

containing this pair, angles (A1 and A2) between the bonds C1'-N9 (of purines) or C1'-N1 (of pyrimidines) and the line passed through both C1' atoms were calculated.

(ii) Deoxynucleotide duplexes. The system consisted of two trinucleotides and one water molecule. At the first step it was determined the low energy B-conformation of (dG)₃:(dC)₃ duplex and then the midpair was substituted by the water bridged pair (C:U or U:U). For calculating the atom coordinates the following variables were taken as independent:

- 1) Six variables determining the position of a trinucleotide as a solid body,
- 2) Six variables determining the position of water molecule,
- 3) 10 variables per nucleotide determining its conformation, 4 of which determined the conformation of the ribose ring. All the bond lengths and all the bond angles (except the angles within sugar ring) were assumed to be constant.

(iii) Codon-anticodon complexes. Codon was represented by the tri-ribonucleotide 5'-CCX-3' and tRNA- by the tetra-ribonucleotide 5'-UYGG-3', corresponding to the 33-36 fragment of the anticodon loop. At first the low energy conformation of the complex CCC UGGG was determined, where CCC had the A-form conformation, and UGGG - the conformation, similar to that of the corresponding fragment of yeast-tRNA^{Phe}. Then the third pair of the codon - anticodon complex was substituted by a pair with water bridge. During the procedure of energy minimization only the conformation of anticodon loop was varied, but the codon always was in the A-RNA conformation. The same independent variables were chosen as for tri-deoxynucleotides and one additional variable per nucleotide, determining orientation of O2'-H-group.

RESULTS AND DISCUSSION

Base pairs with water bridges.

For the systems containing two bases and one or two water molecules at least one minimum with water bridges for every couple of bases have been revealed. The mutual arrangement of the constituents in some of these minima are represented in the Figure 1. The corresponding structural and energetic characteristics are listed in the Table 1. Here we describe the minima, where mutual arrangement of glycosyl bonds of the bases is rather close to that in Watson-Crick pairs. In other publications (ref. 5-7) we represent and discuss the results of calculations for many other such minima. The mutual arrangement of the bases in the minimum 10 (A:U) with water bridge consisted of two water molecules is quite the same as in Watson - Crick

TABLE 1

Energy and structure characteristics of water bridged base pairs^a.

No bases		-Eb	-E1	-E2	R	A1	A2	B1	B2	H1	H2
1	U:C	5.0	5.1	11.7	11.5	32	45	O4	N4H	N3H	N3,O2
		5.6	4.3	0.8						O2	
2	U:U	4.9	5.3	5.3	11.0	56	20	N3H	O4	O4	O4
		0.0	4.9	6.3						O2	N3H
3	C:C	3.0	7.4	8.4	10.2	65	30	N3	N4H	O2	N3
		0.2	3.4	5.7						N4H	N4H
4	A:C	4.1	6.2	7.6	11.9	23	42	N6H	N3	N1	O2
		0.2	10.6	3.5						N7,N6H	N4H
5	G:U	7.1	6.9	4.2	12.1	85	34	N2H	O4	N1H	O4
		0.0	6.3	4.2						N2H	N3H
6	G:A*	5.8	8.4	3.8	10.2	82	57	N2H	N7	N1H	N6H
7	U:A*	4.7	5.1	6.4	12.2	29	25	O4	N6H	N3H	N7
8	G:U	10.6	6.2	5.2	10.6	42	69	N1H,O6	O2,N3H	O6	O4
		0.0	7.4	3.4						N2H	O2
9	^m G:U	4.4	9.5	5.2	10.1	63	88	N2H	O2	N1,O6	N3H
		5.6	0.8	5.0							O4
10	A:U	11.3	5.0	-0.8	10.7	53	55	N6H,N1	O4,N3H	N6H	
		6.1	-0.6	6.1							O4

^aThe second line for every base pair concerns the second water molecule of the complex "2 bases + 2 water molecules". The values of base-base (Eb, first line), water-water (Eb, second line), first base-water (E1) and second base-water (E2) interaction energies are given in kcal/mol. The distance R is given in Å. The angles A1 and A2 are given in degrees. In four last columns the atom groups involved in H-bonding between the bases (B1 and B2) and between the bases and water molecules (H1 and H2) are listed.

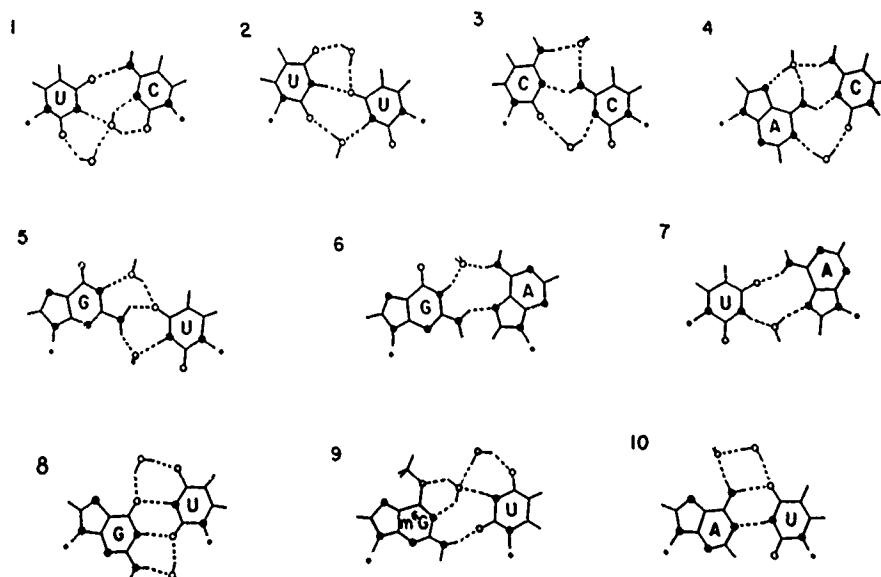


Fig. 1. Water bridged base pairs. Full circles are nitrogen atoms, open ones - oxygens.

pairs. We represent the characteristics of this minimum as the standard for the sake of comparison with those of other minima. In many minima the bases are linked via one H-bond and one or two monomolecular water bridges. In some minima the second water molecule additionally stabilize the configuration by linking with the first water and with one of the bases (e.g. minimum 1). We also have revealed minima with two H-bonds between the bases where water bridges play roles of additional stabilizers of structure. The most bright example of such configuration is G:U wobble pair (minimum 8). The same mutual arrangement of the bases and water molecules has been found in crystals of double helical oligonucleotides d(GGGGTCCC) and d(GGGGCTCC) (ref. 9). The formation of the water bridges is suggested to be one of the reasons of the fact that the G:T (or G:U) pair is the most stable among all "incorrect" pairs.

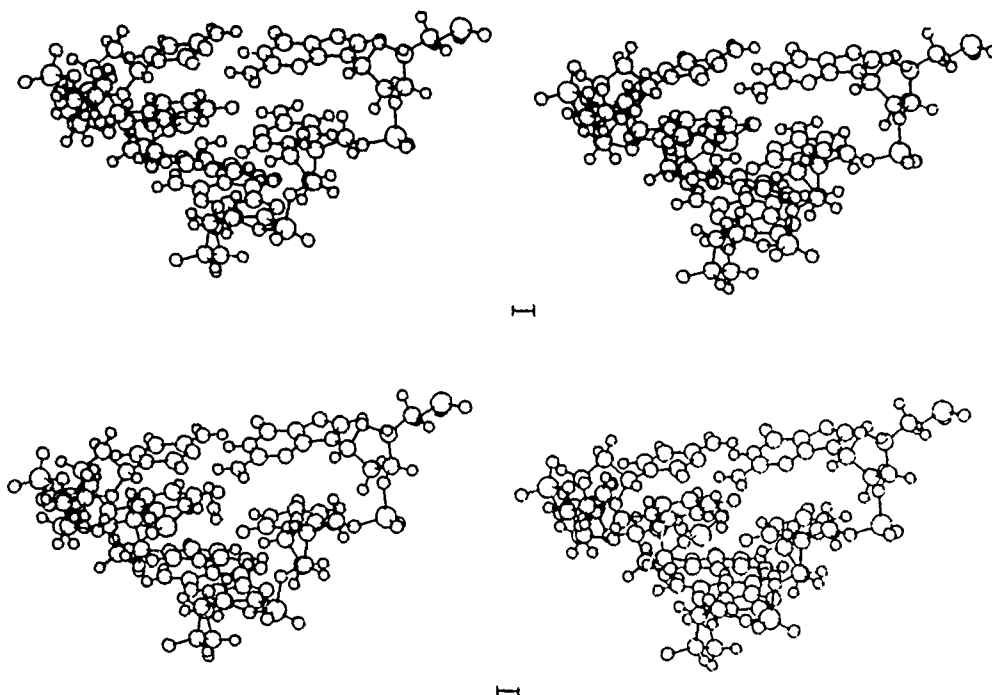


Fig. 2. Stereo views of low energy conformations of the complexes with water bridged mispairs: d(GCG)·d(CUC) (top) and d(GUG)·d(CUC) (bottom). The water molecules forms H-bonds with the bases of U C (U U) mispair and with guanine N3 (cytosine N1) of neighbour pair.

In some minima (e.g. minima 6,7 in the Table 1) one of the bases has syn-orientation with respect to the glycosyl bond.

The mutual arrangement of the glycosyl bonds in some pairs with water

bridges is more close to that in Watson-Crick pairs than in the configuration corresponding to the minimum of energy of base-base interaction. Particularly it concerns to pyrimidine-pyrimidine pairs. We suggest that it is the main reason of a high probability of pyrimidine - pyrimidine pair formation as an error of nucleic acid biosynthesis revealed in several in vitro systems (ref. 10). The NMR data for some oligonucleotide duplexes containing the "incorrect" pairs (ref. 11-13) can be explained by the assumption of the formation of base pairs with water bridges. These pairs correspond to minima 1,4 and 9 in Table 1.

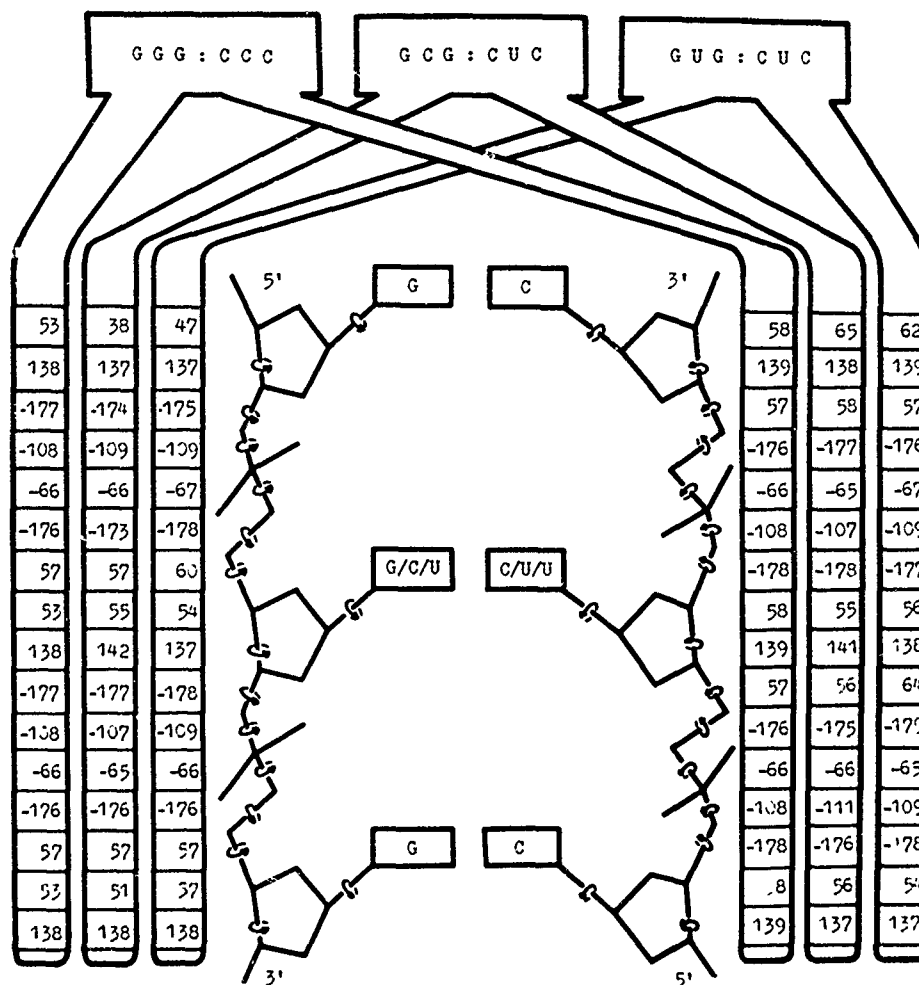


Fig. 3. The comparison of dihedral angle values of the three trinucleotide complexes.

Conformations of deoxynucleotide duplexes.

For the estimating the range of distortions of sugar-phosphate backbone conformation caused by the substitution of Watson - Crick base pair by

mispairs with water bridges we have done the conformational calculations of trinucleotide duplexes, where the flanking positions were occupied by the G:C pairs and the midposition - by the water bridged mispair. Here the results of calculations of d(CUC)·d(GCG) and d(CUC)·d(GUG) complexes are presented (Figs. 2 and 3). The calculations reveal that, in these cases, the deviations of backbone dihedral angles from those of d(CCC)·d(GGG) complex are considerably less than in the cases of mispairs with two H-bonds between the bases (ref. 2,3).

Conformations of codon-anticodon complexes

Analysis of available experimental data suggests a functional asymmetry between the third nucleotide of codon and the first nucleotide of anticodon (ref. 4). For example, C:U pair should be considered as permitted and can be formed in mitochondria, but U:C pair is always impossible. This fact contradicts to Crick's wobble hypothesis, according to which both pairs should be simultaneously either possible or impossible.

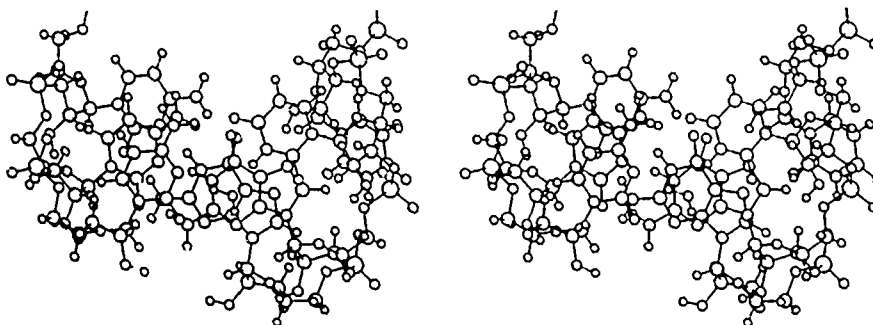


Fig. 4. Stereoview of low energy conformation of the codon-anticodon complex with water bridged C U pair in the third position. The water molecule forms H-bonds with the bases of C U pair and also with guanine N3 of neighbour pair.

To resolve this contradiction it was proposed (ref. 4) that (i) in the codon-anticodon complex the codon is fixed in the A-RNA conformation, and the arrangement of sugar-phosphate backbone, that is necessary for any wobble pair formation occurs only in the anticodon, and that (ii) two bases in the third position of the codon-anticodon complex can mate not only according to the Crick's scheme, but also via water bridges. In such a theory the first of the two pairs, C:U, becomes permitted and the second, U:C, appears sterically impossible. The calculations reveal, that incorporation of the water bridged pair in the third position of the codon-anticodon complex (Fig.4) alters the energy of anticodon loop for less than 1 kcal/mol. That is why such distortions of the loop should be regarded as quiet possible. Analysis of possibility of incorporation of the pairs

with water bridges in the third position of codon-anticodon complex has allowed not only to explain what the pair are possible and what are not, but also to predict the relative efficiencies of their formation.

REFERENCES

- 1 V.I. Poltev and V.I. Bruskov, On molecular mechanisms of nucleic acid synthesis, fidelity aspects: 1. Contribution of base interactions, *J.Theor. Biol.*, 70 (1978) 69-83.
- 2 V.I. Poltev and V.P. Chuprina, Relation of macromolecular structure and dynamics of DNA to the mechanisms of fidelity and errors of nucleic acid biosynthesis, in: E. Clementi et al. (Eds.), *Structure & Motion: Membranes, Nucleic Acids & Proteins*, Adenine Press, New York, 1985, pp. 433-460.
- 3 V.P. Chuprina and V.I. Poltev, Alteration of the DNA double helix conformation upon incorporation of mispairs as revealed by energy computations and pathways of point mutations, *Nucl. Acid. Res.*, 13 (1985) 141-154.
- 4 S.V. Steinberg, Stereochemical model of codon-anticodon interaction with the mating by means of water bridges, *Mol. Biol. (USSR)* 21 (1987) 1322-1328.
- 5 V.I. Poltev and S.V. Steinberg, Interaction of nucleic acid bases via water molecules and formation of mismatched nucleotide pairs, *Mol. Biol. (USSR)*, 21 (1987) 704-713, *Engl. Transl. Molecular biology*, 21 (1987) 588-597.
- 6 V.I. Poltev and S.V. Steinberg, The role of structural water in the formation of nucleotide mispairs, *J. Biomol. Struct. Dyn.*, 2 (1987) 307-312.
- 7 S.V. Steinberg and V.I. Poltev, Water bridged mispairs and the ambiguity of codon-anticodon recognition, *J. Theor. Biol.*, to be submitted.
- 8 V.I. Poltev and N.V. Shulyupina, Simulation of interactions between nucleic acid bases by refined atom-atom potential functions, *J. Biomol. Struct. Dyn.*, 3 (1986) 739-765.
- 9 O. Kenned, Structural studies of DNA fragments: The G.T wobble base pair in A, B and Z DNA; The G A base pair in B-DNA, *J. Biomol. Struct. Dyn.*, 3 (1985) 205-226.
- 10 L.A. Loeb and T.A. Kunkel, Fidelity of DNA synthesis, *Ann. Rev. Bioch.*, 52 (1982) 429-457.
- 11 D.J. Patel, S.A. Kozlowski, S. Ikuta and K. Itakura, Dynamics of DNA duplexes containing internal G T, G A, A C and T C pairs: hydrogen exchange at and adjacent mismatch sites, *Fed. Proc.*, 43 (1984) 2663-2672.
- 12 D.J. Patel and L. Shaper, S.A. Kozlowski, B.L. Gaffney, R.A. Jones, Structural studies of the O6meG T interaction in the d(C-G-T-G-A-A-T-T-C-O6meG-C-G) duplex, *Biochemistry*, 25 (1986) 1036-1042.
- 13 M.W. Kalnik, M. Kouchakdjian, B.F.L. Li, P.F. Swan, D.J. Patel, Base pair mismatches and carcinogen-modified bases in DNA: an NMR study of A C and A O4meT pairing in dodecanucleotide duplexes, *Biochemistry*, 27 (1988) 100-108.

STUDY ON THE SEQUENCE SELECTIVITY OF THE BI-FUNCTIONAL INTERCALATOR DITERCALINIUM

R.C. MAROUN* and B.P. ROQUES

Laboratoire des conformations et des interactions de modèles biologiques et
pharmacologiques (UA498 CNRS) and Unité de Pharmacochimie Moléculaire (U266
INSERM), 4, av. de l'Observatoire, 75006 Paris, FRANCE

SUMMARY

Ditercalinium is a synthetic antitumoral and cytotoxic DNA bi-functional intercalator containing a rigid chain that links its two pyridocarbazole aromatic moieties. Previous theoretical and experimental studies have shown that the complex in which the linking chain of the ligand is in the major groove is favored over the minor groove complex for the case of CG sequences. It was interesting thus to find out whether this was the case for other sequences. In this work the structures and binding energetics of the complexes formed between ditercalinium and a series of model tetradexynucleotide duplexes have been investigated by means of a molecular mechanics procedure especially developed for nucleic acid structures.

A comparative energy analysis of all the complexes has permitted us to order them according to their relative stabilities. The energy minimisation calculations point to an increased stabilisation of the minor groove complex over the corresponding major groove complex in the cases of, for example, d(TACG)₂, d(CATA)₂, and (CATG)₂. Other complexes, such as d(CACG)₂, d(TACA)₂ and d(CACA)₂ are predominantly of the major groove type. A third class includes d(GCAT)₂, d(ACAT)₂, and d(ACGT)₂, which are indifferent energetically to either groove.

INTRODUCTION

In a series of papers we have attempted so far to understand and characterise the complexes formed by the interaction of certain mono and bis-intercalating substances belonging to the 7H-pyridocarbazole family and a number of model oligonucleotides. These substances show a large antitumoral activity and possess a high affinity for DNA (1,2). This endeavor originated in part by the existence of ligands of this family which intercalate to DNA but which, surprisingly, do not possess any antitumoral activity (2). Our studies, theoretical in nature (3-5), have been backed, in several instances, by experimental NMR measurements performed on the same systems (6-8). In characterising the geometrical and energetical properties of the studied systems we have succeeded so far in finding agreement between theory and experiment. In the complexes studied previously, the oligocyclic aromatic moiety of the intercalating ligand was usually lodged in a dCpdG step. Non-covalent bifunctional intercalation in the case of ditercalinium (denoted here as 202; Fig. 1) and its

derivatives proceeded in all cases via the major groove of the nucleotide, i.e., the linking chain of the drug occupied the major groove side of the oligonucleotide.

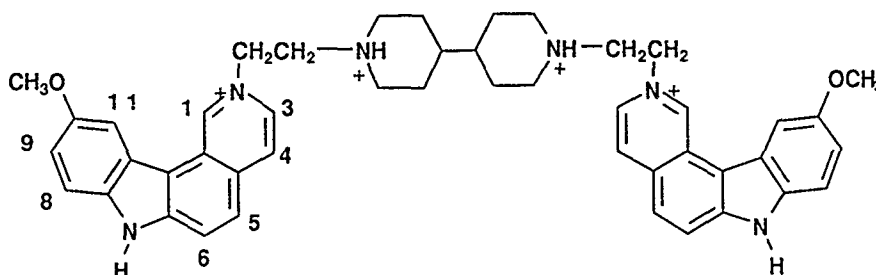


Figure 1 Chemical structure of ditercalinium

We were curious, then, to find out whether intercalation from the major groove side was a rule for this family of ligands. For that purpose, the cytosine and guanine bases are replaced by or combined with thymine and adenine in a rational fashion, always following an alternating (purine-pyrimidine)₂ or (pyrimidine-purine)₂ sequence. In all the complexes the two central base pairs are sandwiched between the chromophores intercalated at the terminal sites. The resulting complexes are studied by energy minimisation.

METHODS

The theoretical methodology utilised in this work is that of the JUMNA molecular mechanics procedure. This technique has been already described in detail (9-10).

RESULTS AND DISCUSSION

Table I shows the DNA-ligand stabilisation energies for the optimised complexes obtained between ditercalinium and a series of model tetradeoxyribonucleotides of alternating pyrimidine-purine or purine-pyrimidine pattern. In these complexes the linking side chain of the ligand is located either in their major groove side (M) or in the minor groove side (m). The methoxy group at position 10 of the drug chromophore is always oriented towards the central base pairs.

TABLE I

Summary of the energy terms for the optimal bis-intercalating complexes of ditercalinium with selected model oligonucleotides (values in kcal/mol)

Nucleotide in complex ¹	groove	ΔE_{DNA}	ΔE_{LIG}	E_{INTER}	E_s	∂E_s^2
1a) CACA	M	59.4	4.0	-165.2	-102.2	-18.5
b)	m	57.6	0.5	-141.8	-83.7	
2a) CACG	M	59.9	0.4	-168.4	-108.1	-22.0
b)	m	57.6	0.4	-144.1	-86.1	
3a) CATA	M	50.7	1.6	-123.8	-71.5	41.8
b)	m	61.4	2.5	-177.2	-113.3	
4a) CATG	M	53.9	1.4	-122.4	-67.1	46.2
b)	m	63.4	2.7	-179.4	-113.3	
5a) CGCA	M	49.7	4.1	-151.7	-97.9	10.2
b)	m	42.6	1.2	-152.1	-108.3	
6a) CGCG	M	67.8	5.6	-185.9	-112.5	-1.3
b)	m	58.8	2.1	-172.1	-111.2	
7a) TACA	M	54.1	3.3	-160.0	-102.6	-15.0
b)	m	58.7	0.5	-146.8	-87.6	
8a) TACG	M	54.5	1.6	-161.9	-105.8	10.1
b)	m	59.9	0.6	-176.4	-115.9	
9a) TATA	M	39.5	4.1	-116.9	-72.3	35.8
b)	m	55.4	7.2	-170.7	-108.1	
10a) TGCA	M	41.4	3.3	-142.0	-97.3	-20.8
b)	m	43.8	2.2	-122.5	-76.5	
11a) ACAC	M	53.9	1.0	-139.2	-84.3	-10.6
b)	m	54.1	0.5	-128.3	-73.7	
12a) ACAT	M	50.1	1.3	-123.5	-72.1	2.8
b)	m	51.7	0.6	-127.2	-74.9	
13a) ACGC	M	57.9	1.2	-131.0	-71.9	-7.5
b)	m	55.8	-0.1	-120.1	-64.4	
14a) ACGT	M	48.3	1.5	-118.5	-68.7	-2.8
b)	m	41.5	-2.7	-104.7	-65.9	
15a) ATAC	M	47.0	1.5	-140.8	-92.3	-18.1
b)	m	44.2	2.0	-120.4	-74.2	
16a) ATAT	M	40.9	1.0	-129.1	-87.2	-8.4
b)	m	33.1	3.9	-115.8	-78.8	

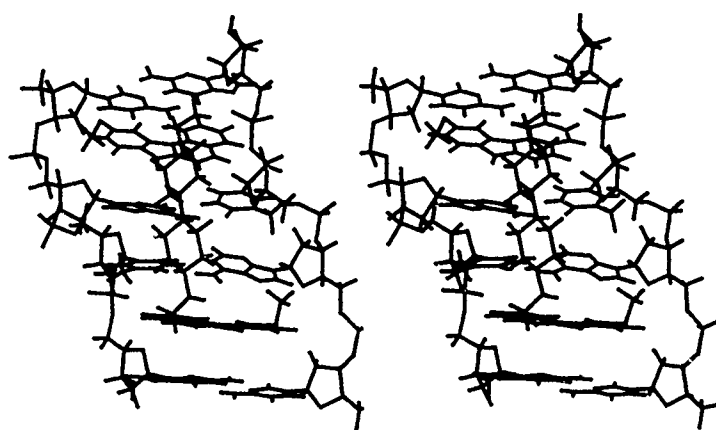
17a) GCAC	M	63.8	0.9	-146.8	-82.1	-13.5
b)	m	60.0	0.7	-129.3	-68.6	
18a) GCAT	M	58.2	1.4	-135.8	-76.2	-1.5
b)	m	60.8	-1.2	-134.3	-74.7	
19a) GCGC	M	70.9	2.4	-151.5	-78.2	-22.0
b)	m	46.1	-1.5	-100.8	-56.2	
20a) GTAC	M	50.0	1.0	-148.0	-97.0	-23.7
b)	m	46.3	2.2	-121.8	-73.3	

¹ Bisintercalation takes place at the end sites BP1-BP2 and BP3-BP4 (BP: base pair).

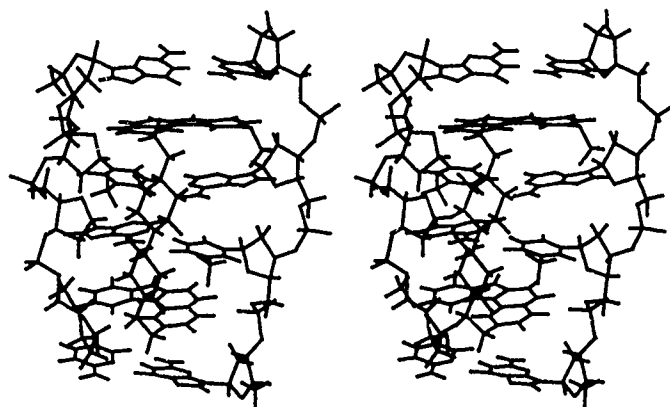
² ∂E_s : Energy changes of major groove with respect to minor groove complex.

³ The total complex energy $E_s = \Delta E_{DNA} + \Delta E_{LIG} + E_{INTER}$, where ΔE_{DNA} is the intramolecular DNA energy change with respect to free DNA; ΔE_{LIG} is the intramolecular ligand energy change with respect to free ligand; and E_{INTER} is the intermolecular DNA-ligand energy.

From all the 20 possible pairs of alternating (Py Pu)₂ or (Py Pu)₂ tetranucleotides, ditercalinium bisintercalates from the major groove side (M) in 11 of them, from the minor groove side (m) in 5 of them, and indifferently (within 3 kcal. of ∂E_s) in 4 of them. Figure 2 shows the molecular graphics representations of complexes 4b and 20a.



a) 202/CATG (m)



b) 202/GTAC (M)

Figure 2. 'Relaxed' stereo view from major groove side of a) 202 in minor groove of CATG ; b) 202 in major groove of GTAC.

Table II shows the outcome if the more favorable complex, major or minor groove, from each pair of complexes in Table I is chosen and then arranged according to decreasing stabilisation.

TABLE II

Arrangement in order of decreasing stabilisation of the most favored -minor groove or major groove- of each pair of the complexes in Table I

NUCLEOTIDE	PREFERRED GROOVE	SEQUENCE PATTERN
TACG	m	(PyPu) ₂
CATA	m	(PyPu) ₂
CATG	m	(PyPu) ₂
CGCG	M,m	(PyPu) ₂
CGCA	m	(PyPu) ₂
CACG	M	(PyPu) ₂
TATA	m	(PyPu) ₂
TACA	M	(PyPu) ₂
CACA	M	(PyPu) ₂
TGCA	M	(PyPu) ₂
GTAC	M	(PuPy) ₂
ATAC	M	(PuPy) ₂
ATAT	M	(PuPy) ₂
ACAC	M	(PuPy) ₂
GCAC	M	(PuPy) ₂
GCGC	M	(PuPy) ₂
GCAT	M,m	(PuPy) ₂
ACAT	m,M	(PuPy) ₂
ACGC	M	(PuPy) ₂
ACGT	M,m	(PuPy) ₂

We can observe from this table II that the most favorable complexes are those i) following the (Py Pu)₂ alternating sequence, and ii) mainly forming minor groove bis-intercalation; in other words, the minor groove complexes are in general more stable than the major groove ones.

Table III is obtained if for a given groove the complexes are arranged by energy differences with respect to complex 6a (dEc).

TABLE III

Arrangement of the complexes of Table I in order of decreasing destabilisation with respect to 202/d(CGCG)₂

MAJOR GROOVE		MINOR GROOVE	
NUCLEOTIDE	dEc	NUCLEOTIDE	dEc
CGCG	0.0	TACG	-3.4
CACG	4.4	CATA	-0.8
TACG	6.7	CATG	-0.8

TACA	9.9	CGCG	1.3
CACA	10.3	CGCA	4.2
CGCA	14.6	TATA	4.4
TGCA	15.2	TACA	24.9
GTAC	15.5	CACG	26.4
ATAC	20.2	CACA	28.8
ATAT	25.3	ATAT	33.7
ACAC	28.2	TGCA	36.0
GCAC	30.4	ACAT	37.6
GCGC	34.3	GCAT	37.8
GCAT	36.3	ATAC	38.3
TATA	40.2	ACAC	38.8
ACAT	40.4	GTAC	39.2
ACGC	40.6	GCAC	43.9
CATA	41.0	ACGT	46.6
ACGT	43.8	ACGC	48.1
CATG	45.4	GCGC	56.3

We can observe from this table III that i) for either mode of bis-intercalation, major groove or minor groove, the (PyPu)₂ pattern is favored over the (PuPy)₂ one; ii) the tetramers TACG, CGCA and CGCG form stable complexes for both grooves; ACGT and ACGC form the least stable complexes for both grooves; iii) a weak preference for CpA, CpG, TpA, and TpG is observed, based on the number of times a given step appears favored within 15 kcal of dEc; iv) aside from (i) no other trend for a coupled dinucleotide BP₁-p-BP₂ to BP₃-p-BP₄ step preference can be extracted neither from the major nor from the minor groove, nor when comparing both grooves. (BP: base pair).

Ditercalinium then does not seem to show any particular strong sequence specificity. On the other hand, depending on the sequence, it possesses a certain groove specificity. This is what we call **sequence-dependent groove preference**.

CONCLUSION

The energy minimisation calculations indicate:

- an increased stabilisation of the minor groove complex over the corresponding major groove complex in the case of TACG, CATA, CATG, CGCA, and TATA. Other complexes such as the one with CACG and TACA are major groove complexes. Complexes with CGCG, GCAT, ACAT, and ACGT are indifferent energetically to either groove.
- the number of favored major groove complexes is larger than the number of minor groove ones.
- regardless of the groove to which ditercalinium bisintercalates, the (PyPu)₂ pattern is always more favored than the (PuPy)₂ one.

Finally, the prediction of bis-intercalation from the minor groove side in the case of the TATA sequence is in agreement with the theory of the molecular electrostatic

potential as applied to DNA (11, and references therein). The mode of intercalation of ditercalinium with a TpApTpA sequence is presently being investigated in this same laboratory by ^1H and ^{31}P NMR techniques.

REFERENCES

1. D. Pelaprat, A. Delbarre, I. Le Guen, B. P. Roques and J. B. Le Pecq, **J. Med. Chem.**, 23 (1980) 1336-1343.
2. C. Garbay-Jaureguierry, P. Laugâa, M. Delepierre, S. Laalami, G. Muzard, J. B. Le Pecq, and B. P. Roques, **Anti-Cancer Drug Design**, 1 (1987) 323-335.
3. R. Maroun and N. Gresh, **Biopolymers**, 4 (1989) 835-849.
4. R. Maroun, M. Delepierre and B. P. Roques, **J. Biomol. Struct. and Dyn**, 3 (1989) 1-15.
5. R. Maroun, M. Delepierre and B. P. Roques, in preparation.
6. A. Delbarre, M. Delepierre, C. Garbay, J. Igolen, J. B. LePecq and B. P. Roques, **Proc. Natl. Acad. Sci USA**, 84 (1987) 2155-2159.
7. A. Delbarre, M. Delepierre, J. Langlois d'Estaintot, J. Igoien and B. P. Roques, **Biopolymers**, 26 (1987) 1001-1033.
8. M. Delepierre, R. Maroun, C. Garbay-Jaureguierry, J. Igolen and B. P. Roques, **J. Mol. Biol.**, (1989) in press.
9. R. Lavery in: W.K.Olson, M.Sundaralingam, M.H.Sarma and R.H.Sarma (Eds.), **"DNA Bending and Curvature,"** Adenine Press, New York, NY, 1988, pp. 191-201.
10. R. Lavery in: R.D.Wells and S.C.Harvey (Eds.), **"Unusual DNA Structures,"** Springer Verlag, 1988, pp. 189-206.
11. K. Zakrzewska, R. Lavery and B. Pullman, **Nucl. Acids Res.**, 12 (1984) 6559.

NMR STUDIES OF PROTEINS AND PROTEIN-DNA INTERACTIONS

J.A.C. Rullmann, R. Boelens, R.M.J.N. Lamerichs, G.W. Vuister and R. Kaptein
Department of Chemistry, University of Utrecht, Padualaan 8,
3584 CH Utrecht (the Netherlands)

SUMMARY

An overview is given of recent NMR studies on the structures of crambin, *lac* repressor headpiece and the complex formed between the latter and *lac* operator. The solution structure of crambin was determined using relaxation matrix calculations. Aspects of this technique are discussed. Docking calculations for the *lac* headpiece-operator complex are reviewed. The structural consequences of new NMR data obtained for *lac* headpiece are discussed. Advantages and applications of three-dimensional NMR spectroscopy are addressed.

INTRODUCTION

In the last decade NMR spectroscopy has emerged as an important tool for structure determination of biomolecules in solution. The availability of high-field spectrometers and fast computer systems made it possible to develop two-dimensional (2D) NMR techniques by which the resonance assignment problem could be tackled (ref. 1). Strategies to assign ^1H proton resonances in protein and DNA spectra were developed (ref. 2). Finally methods have become available to determine the structure on the basis of distance constraints from NMR, such as Distance Geometry (DG) and restrained Molecular Dynamics (MD) (refs. 2, 3).

NMR structures are primarily based on proton-proton distances obtained from the nuclear Overhauser effect (NOE). The origin of the NOE is dipolar cross-relaxation between protons. In first order the strength of the NOE is inversely proportional to the sixth power of the distance between two protons. By calibrating against NOEs observed for protons at a known distance relative cross-peak intensities can be translated into constraints on proton-proton distances, which can then be used to determine the structure. An important drawback of this so-called two-spin approximation is that it ignores spin diffusion effects i.e. contributions to dipolar relaxation via other neighbouring protons. This approximation has been shown to yield distances that may be in error by as much as 20 to 40 percent (ref. 4). In order to account for spin diffusion it is necessary to solve the relaxation equations for all spins simultaneously. A new method to handle this problem will be discussed.

Furthermore we will review the structural studies performed with NMR of the interactions between a protein, the *lac* repressor headpiece, and its DNA binding site. These studies are a good example of the power of NMR methods to elucidate important structural details. Not only was the orientation of the protein with respect to

the DNA found to be approximately 180° different from that of other DNA binding proteins (ref. 5), but from the NMR structure also specific protein-DNA interactions were predicted which have been found recently in genetic experiments with the complete *lac* repressor (ref. 6).

Finally we will discuss the development of three-dimensional (3D) NMR experiments. One of the advantages of the new technique is the increased resolution, offering the possibility of analyzing larger molecules. With 3D-techniques it also becomes possible to directly observe spin diffusion pathways.

RELAXATION THEORY

Dipolar relaxation in a multispin system is approximately described by a Bloch type equation of the form

$$A(t_m) = A(0) \exp(-R t_m) \quad (1)$$

where A is the matrix of normalized cross-peak intensities observed in a 2D NOE experiment with mixing time t_m . R is the relaxation matrix; its elements R_{ij} can be related to the spectral moments of the dynamical behaviour of the position vector connecting spin i to spin j . Assuming that the molecule is rotating isotropically while remaining rigid, the relaxation matrix elements may be calculated theoretically from the interproton distances and the rotation correlation time t_c (refs. 1, 7, 8):

$$R_{ij} = \frac{K}{d_{ij}^6} \left(\frac{6 t_c}{1 + 4 \omega^2 t_c^2} - t_c \right), \quad R_{ii} = \sum_{j \neq i} \frac{K}{d_{ij}^6} \left(\frac{6 t_c}{1 + 4 \omega^2 t_c^2} + \frac{3 t_c}{1 + \omega^2 t_c^2} + t_c \right) \quad (2)$$

where $K = 0.1 \gamma^4 h^2 (2\pi\mu_0)^{-2}$. For short mixing times eqn. (1) may be expanded in a Taylor series

$$A(t_m) = A(0) \{ 1 - R t_m + R^2 t_m^2 / 2 - \dots \} \quad (3)$$

Keeping only the first two terms one simply has

$$A_{ij} \propto t_m t_c d_{ij}^{-6} \quad (4)$$

When the distance for a particular proton pair is known, the corresponding NOE intensity A_{cal} can be used to convert intensities to distances, using

$$d_{ij} = d_{cal} (A_{cal} / A_{ij})^{1/6} \quad (5)$$

In practice it is difficult to obtain accurate NOE values at mixing times that are sufficiently short for the approximation of eqn. (4) to be valid. At longer mixing times the NOEs are stronger and more easily observed, but the two-spin approximation then breaks down and one observes indirect magnetization transfer or "spin diffusion": neighbouring spins start to contribute to the buildup of NOE intensity. For this reason one usually translates the NOE information into distance *ranges* rather than precise

distances (refs. 2, 3).

RELAXATION MATRIX CALCULATIONS

Iterative Relaxation Matrix Approach

It is possible to treat spin diffusion exactly within the context of eqn. (1). By diagonalization of the complete relaxation matrix for a given structural model eqn. (1) may be solved to calculate the NOE matrix which can then be compared with the experimental data (ref. 8). This method has been used to look at various model structures of oligonucleotides in solution (ref. 9). Working in the opposite direction it is possible to "back transform" the experimental NOE matrix to obtain the cross-relaxation rates and thus the interproton distances corrected for spin diffusion (ref. 10).

For medium-sized and large biomolecules like proteins and DNA only part of the NOE matrix can be obtained experimentally, and the direct computation of R from eqn. (1) is therefore not possible. As first shown by Boelens et al. (refs. 4, 11) this problem can be circumvented by supplementing the experimental information with theoretical data calculated from a model structure. The so-called Iterative Relaxation Matrix Approach (IRMA) starts from an initial structure, e.g. a linear chain. Theoretical NOE intensities are computed and combined with the experimental values after scaling of the latter using a suitable set of calibration peaks, i.e. theoretical values are replaced by experimental ones for each proton pair for which a NOE has been observed. Using eqn. (1) a new relaxation matrix is computed from the combined NOE matrix. The off-diagonal relaxation matrix elements R_{ij} are directly translated into distance bounds, taking into account the uncertainty in R_{ij} as reflected in its variation with the mixing time (ref. 11). Structure calculations (DG and/or MD) are then performed with the calculated distance bounds on all proton pairs for which NOEs were observed. The resulting structure is used as starting point for a new IRMA cycle, i.e. calculation of theoretical NOEs etc. The process is repeated until the structure and the distance constraints converge.

IRMA was first applied to a DNA octamer (ref. 11) and showed good convergence properties, irrespective of the starting structure (both A- and B-DNA models were used). The accuracy of the procedure was such that effects of local mobility, which are not included in the present treatment, could be observed in the pattern of residual constraint violations.

The importance of correcting for spin diffusion is shown by a comparison of distances as calculated directly from theoretical NOE buildups via the two-spin approximation, eqns. (4-5), with the real distances (ref. 4). The relative error varies linearly with the distance; distances not corrected for spin diffusion have a narrower distribution than the real distances. For the calibration distance used in this example the error ranges from -20 to +20 percent.

Results for crambin

Recently IRMA has also been applied to a protein (ref. 12). The solution structure of crambin was determined using more than 600 experimental NOEs, each measured at 6 different mixing times. Effects of methyl rotation and aromatic ring flips were taken into account (ref. 13). Structure calculations employed the Iterative Relaxation Matrix Approach in combination with DG and MD simulations. The initial input to IRMA, used to calculate the first set of constraints, consisted of the fully extended peptide chain. The final structures after two cycles of IRMA show a backbone r.m.s. variation of approximately 1 Å. The difference with the X-ray structure is also 1 Å (ref. 12).

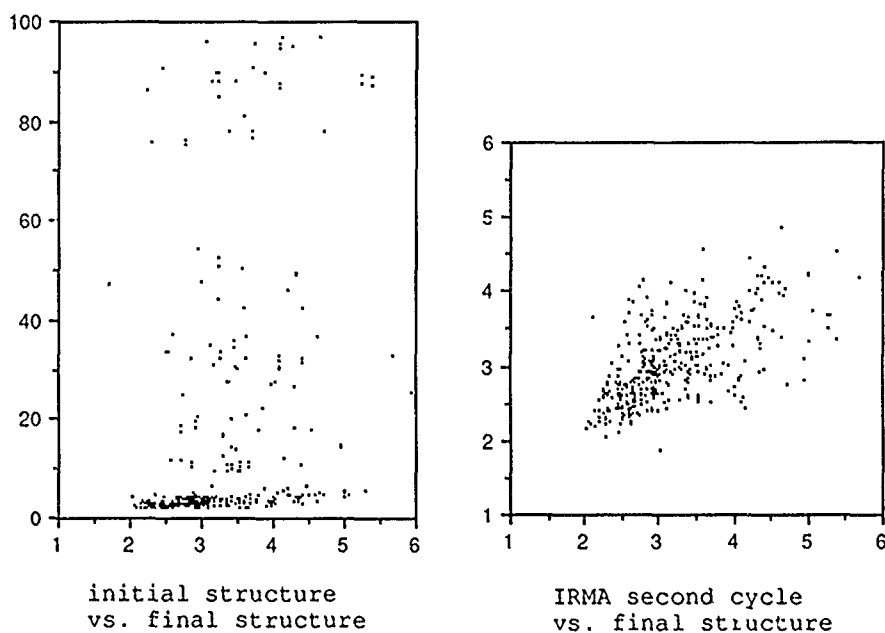


Fig. 1a (left) and 1b (right). Diagrams showing convergence of IRMA procedure. For each observed NOE a point is plotted. Horizontal axis: distance in final structure. Vertical axis: distance in initial structure (left) and distance from IRMA after two cycles (right).

Fig. 1 illustrates how IRMA succeeds in transforming the experimental data into distance constraints. Fig. 1a plots distances d_{ext} in the initial extended chain structure against distances d_{sol} in one of the best final solution structures; only data corresponding to observed NOEs are shown. The r.m.s. difference $\langle (d_{\text{ext}} - d_{\text{sol}})^2 \rangle^{1/2}$ is 24.8 Å. The distribution of distances after two cycles is shown in Fig. 1b. Due to the large number of experimental data the calculations converge rapidly. Already in the

first cycle, before any structure calculation, the r.m.s. difference between IRMA distances d_{IRMA} and d_{sol} is down to 0.58 Å. This value is virtually unchanged in the next cycles. The r.m.s. in terms of IRMA upper bounds is even lower: 0.43 Å.

It may be somewhat surprising that the final structure does not completely satisfy the IRMA upper bounds. This is caused by the impossibility to discriminate experimentally between protons on prochiral centres, such as β - β' pairs and the two methyl groups of Val and of Leu residues. In the plots we simply calculated the distance to the average positions in these cases. In the simulations, however, pseudo atoms were used and many bounds were increased with 1 Å or more, which reduces the precision with which side chains can be positioned. So although the optimized structures in the simulations all have low average constraint violations of less than 0.1 Å, which seems well within the present accuracy of the IRMA method, the theoretical buildups predicted from these structures may not completely match the experimental NOEs.

Using an NMR *R*-factor as measure for the correspondence between calculated and measured NOEs it is possible to discriminate between structures, and also pinpoint regions or groups in the molecule for which the correspondence is less good (ref. 12). In the same spirit it appears possible to obtain stereospecific assignments directly from a comparison between theoretical and experimental NOE values (data not shown). Together with methods to calculate local mobility corrections to the rigid molecule model these procedures provide the framework for extending the applicability of IRMA in obtaining high-quality structures with NMR.

LAC REPRESSOR HEADPIECE - OPERATOR COMPLEX

Main structural features

One of the largest systems studied by NMR is the specific protein-DNA complex formed between the headpiece, i.e. the N-terminal fragment, of *lac* repressor protein and a symmetric 22 base pair (bp) *lac* operator fragment (ref. 14). The biological function of *lac* repressor is to regulate the transcription of the *lac* genes in *E. coli*. Expression is shut off by binding of repressor to the operator; transcription of the lactose enzymes can be induced again by allolactose which causes the protein to lose its specific binding capacity and thus disrupts the complex. The system has been very well characterized, both biochemically and genetically (ref. 15). The natural repressor binds as a tetramer to DNA. The headpieces of two subunits appear to be involved in binding to the operator, which indeed has an approximate two-fold symmetry. Proteolytic cleavage yields headpiece fragments consisting of 51, 56 or 59 residues, depending on the proteolytic enzyme. These fragments retain their original 3D structure and their ability to bind specifically to a half-operator.

The complete *lac* repressor, with a molecular weight of 154,000, is too large to study in detail with present day NMR techniques. As model systems we therefore employed the complexes formed between different headpieces (51, 56, 59) and

synthetic operators (11 or 14 bp) corresponding to the left, i.e. the more strongly binding, half of the wild type operator (ref. 15). Recently we also analysed the NOE spectra of the complex formed between the 22 bp symmetric operator fragment and two headpieces, which has a molecular weight of approximately 25,000 (ref. 14).

First the structure of headpiece (HP) itself was determined, using a combination of model building and restrained MD simulations (refs. 16, 17). The molecule was found to consist of three α -helices packed around a hydrophobic core. The DNA binding domain appeared to fold in a helix-turn-helix motif, confirming suggestions based upon sequence homology between the DNA binding domains of *lac* repressor and a number of other proteins for which crystal structures had been determined, such as *cro*, λ and *trp* repressors, and CAP (ref. 15, 18).

From 2D NMR studies on the *lac* headpiece complex, first performed with HP56 - 14bp (refs. 5, 19) and later with HP56 - 11bp and HP56 - 22bp combinations (refs. 14, 20), it readily became apparent that both protein and DNA essentially retain their free-state conformations. The packing of the helices stays intact, and the operator remains in its B-DNA type structure, although some local deformations can not be excluded. No essential differences are observed between the different complexes, apart from an apparent difference in binding strength.

Despite the complexity of the 2D spectra due to the size of the system, which has a molecular weight of 14,000, a number of NOEs between protein and DNA could be identified. These involve only contacts between apolar groups, since the H₂O spectra in which exchangeable protons can be observed are difficult to analyse in detail. Nevertheless the pattern of NOEs is consistent with the data obtained in protection experiments and in genetic studies: the second or "recognition" helix, extending from residue 17 to 25, binds in the major groove of DNA.

This mode of interaction corresponds to what had been inferred from genetic data and model building studies, both for *lac* headpiece and for other repressors (ref. 18). However, the NMR data show conclusively that the orientation of the *lac* headpiece recognition helix with respect to the DNA axis is the opposite of that in the model proposed on the basis of the supposed analogy to the models of *cro* and CAP complexes.

As is clear from Fig. 2 the cluster of NOE contacts seen from Thr-5, Leu-6 and Tyr-7 in the first helix to base pairs 9 and 10, and the NOE cluster from His-29 in the loop between the second and the third helix to base pairs 2 and 3, uniquely determine the orientation of the protein. Further evidence comes from an NMR study of the complex between the 22 bp symmetric operator and headpiece 56 (ref. 14). This 2:1 complex was found to be completely symmetric, which means that the two headpieces bind simultaneously in identical fashion. The NOE pattern showed that the binding mode is the same as that observed in complexes of one headpiece with a half operator.

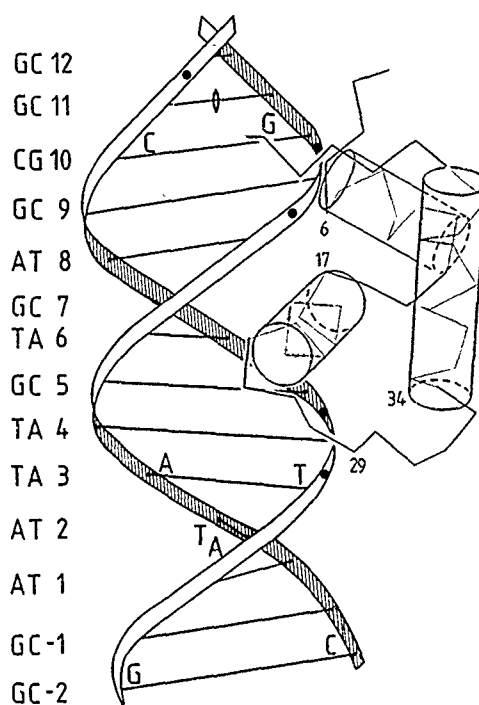


Fig.2. Schematic drawing of the *lac* headpiece operator complex. The black dots indicate the phosphates where ethylation interferes with *lac* repressor binding (ref. 37).

These results support the idea that the complete repressor binds with the same "reversed" orientation as the monomeric model systems. Recent genetic experiments have provided evidence that this is indeed the case. Lehming et al. (ref. 6) investigated the binding between operators and repressors which had mutations at one or two positions. By correlating replacements with affinity changes they inferred interactions between the first two residues of the recognition helix and base pairs 7 and 9, and between the sixth residue (Arg-22 in wild type) and base pair 5. Especially the latter result is strong evidence that the orientation observed with NMR is the same as that of the complete repressor.

By now a number of crystal structures of repressor-operator complexes have been determined (refs. 21-23). In all these cases the orientation agrees with the earlier models proposed for λ , *cro* etc. Therefore two modes of binding of helix-turn-helix proteins to DNA appear to exist, which can be designated as the *lac* and *cro* types. From homology arguments and the genetic data obtained by Lehming et al. (ref. 24) it is likely that at least *gal* and *deo* repressors and maybe also CAP (ref. 6) belong to the *lac* type binding mode.

Detailed simulations

On the basis of the NMR data we attempted to model the structure of the *lac* headpiece-operator complex in more detail. The uncertainties here are much larger than for headpiece itself. Firstly the number of NOEs observed between protein and DNA is quite small (24 at the time of these simulations). Secondly, these NOEs were determined at relatively long mixing times in order to enhance their appearance. Therefore spin diffusion contributes to the intensity, and relatively large upper distance bounds had to be used (but this can be remedied by using IRMA). We used limits of 4 and 6 Å depending on the strength of the NOE.

An important problem in such studies of an intermolecular complex is how an optimal alignment of the two subunits can be attained. Visual inspection using a molecular graphics system obviously has its limitations; on the other hand restrained MD simulations are liable to end up in the wrong minimum, since the molecules do not easily disengage on a picosecond time scale once they are locked into a particular configuration. We obtained good results with another kind of optimization procedure called the Ellipsoid Algorithm. Essentially this is a minimization procedure which avoids local minima by making large, discontinuous steps of gradually decreasing size. An extra advantage stems from handling only one constraint per step instead of the total constraint function.

Using Billeter's implementation (ref. 25) we attempted the docking of rigid headpiece and operator fragments (ref. 26). Initial co-ordinates were taken from a model built structure of the complex, which had been optimized by a very short (5 ps) MD run (ref. 27). The total violation of protein-DNA constraints was less than 1 Å after each of the three dockings that were performed. By comparison: the total violation was 3.6 Å in the starting structure and 13.0 Å in the model built one. In the docking procedure the headpiece moved along the DNA by almost one base pair, corresponding to an r.m.s. change in C_{α} positions of 5 to 7 Å.

While the ellipsoid algorithm is good in finding global minima of a relatively smooth potential, such as the total protein-DNA constraint function, it is less good at optimizing local details such as the exact position of side chains, which involve rapidly changing repulsive potentials. Further optimizations were therefore performed with restrained EM and MD using the GROMOS package (ref. 28). As before the DNA was kept in the B-DNA form, and the constraints as derived for the free headpiece were applied to the protein.

In these runs we also put hydrogen bond constraints between G7 and Gln-18 and between G5 and Arg-22. The contact between G7 and Gln-18 was assigned on the basis of mutant studies by Ebright (ref. 29) and by Lehming et al. (ref. 24). The contact between G5 and Arg-22 was postulated by us on the basis of the NMR data and model building (ref. 5) and is supported by recent mutant studies (ref. 6). Because it was not known a priori which nitrogen acts as hydrogen bond donor in Arg and whether N or O is the acceptor atom in Guanine, the constraints had to be fairly liberal: 6.0 Å for Gln

and 7.3 Å for Arg.

Several MD simulations of 10 or 20 ps were performed starting from different configurations (ref. 26). The total constraint violations always dropped rapidly to a few 0.1 Å. Resulting structures had C_{α} r.m.s. differences up to 3.5 Å. One of the best structures is shown in Fig. 3. Surprisingly few hydrogen bonds between protein and DNA were found. The most persistent ones are those between the amide of Leu-6 and the phosphate of C9, and between N_{ϵ} of Gln-18 and (mainly) N7 of G7. The latter result shows that one of the two postulated contacts is perfectly compatible with the NMR data.

The Arg-22 side chain on the contrary almost never formed a hydrogen bond with G5. Instead contacts were sometimes observed with T3 and T4. Since the non-NMR constraints applied to test the formation of a hydrogen bond were liberal we do not think that the presence of these extra constraints has seriously influenced the calculated structure. The only other group in the recognition helix that formed hydrogen bonds to bases of DNA was Ser-21, which was found to interact alternatingly with three different bases, G7, C7 and T8.

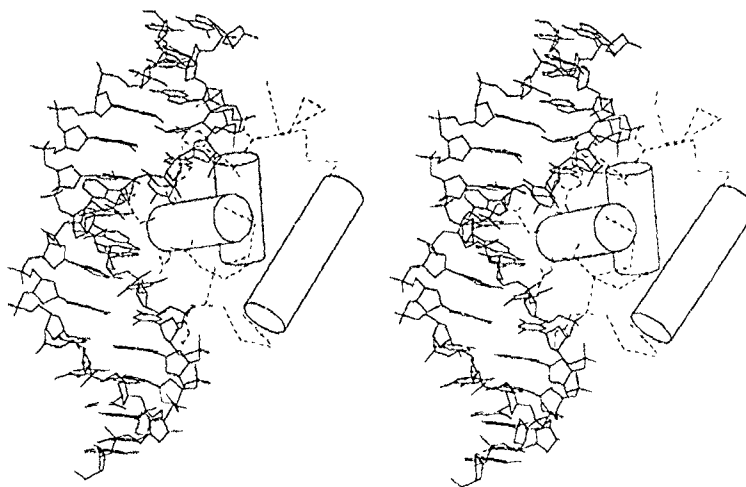


Fig. 3. Stereoview of a structure of the *lac* headpiece - operator complex, obtained by ellipsoid docking, followed by restrained MD and EM (ref. 26).

A similar picture has come forward from simulations of the *lac* headpiece - operator complex in water, performed by de Vlieg et al. (ref. 30). In a trajectory of 125 ps, partly simulated without the NMR constraints, few hydrogen bonds were found between the protein side chains and DNA bases. The ones that did occur were mediated by a water

molecule. On the other hand various long lived non-polar contacts were seen. This raises the question what (if any) the contribution of hydrogen bond interactions to specific recognition can be.

New data

In the last year the identification of NOE cross peaks in spectra of the complex has been extended, partly by measuring at a spectrometer frequency of 600 instead of 500 MHz, and partly by analysing spectra of the HP56 - 11 bp variant, which show less overlap than spectra of the other systems. The total number of constraints for the protein has increased significantly from 137 to 207. Subsequently many NOEs could be traced back in the spectra of the free headpiece; however the constraint sets for the free and the bound form are no longer identical.

We performed a preliminary study of the structural consequences of the new constraint data. Starting from a previously obtained MD configuration 80 ps of restrained MD were carried out, both for the "free" and the "bound" states, of which the last 40 ps were used for analysis. Averaged and subsequently minimized structures appear to differ significantly from the previous model. Similarly a difference is found between the free and bound structures, which is localized around His-29 in the loop connecting helix 2 to helix 3.

The NOE analysis uncovered also a number of new protein-DNA contacts. In total 40 of these have now been identified, while 24 were used in the simulations discussed above. New contacts involve His-29, Gln-18 as well as Tyr-7 and Tyr-17. In the $^1\text{H}_2\text{O}$ spectrum a NOE cross peak has been identified as a contact between the side chain amide of Gln-18 and the H5 proton of base C7. This is in agreement with genetic data (refs. 24, 29) but indicates that the Gln side chain is interacting with C7 rather than G7. Preliminary modelling results show that this contact can easily be accommodated. New structure calculations are presently being undertaken, and we expect that the new NOE data, in combination with the use of IRMA (ref. 11), will increase the precision of the structure. An important piece of information may come from the NOEs observed for Arg-22: contacts to the protein backbone now appear to constrain the side chain to a more extended conformation than before, which brings the guanidyl group in a much closer contact with the G5 base. In what way this will affect the hydrogen bonding pattern remains to be seen.

THREE-DIMENSIONAL NMR SPECTROSCOPY

2D NMR has allowed the detailed analysis of NOE and J-interactions in molecules up to a molecular weight of approximately 15,000 (ref. 2). Beyond this limit the increase in line width and the number of resonances cause too much overlap, even in 2D spectra. Specific isotope labeling or comparison of spectra obtained in different experimental conditions may be used to overcome this problem. A much more general method for studying complex systems is the recently developed 3D NMR technique

(refs. 31-33). The addition of a third frequency domain helps in two ways. First, the spectra are less crowded since the interactions, in the form of cross peaks, are spread out over a larger area. Secondly, assignment of interactions that overlap in 2D spectra becomes possible in 3D when the additional coherence transfer step links them to different uniquely defined nuclei.

Until now several homonuclear and heteronuclear 3D NMR techniques have been proposed, such as homonuclear 3D NOE-HOHAHA (refs. 32, 33), 3D NOE-NOE (ref. 34) and heteronuclear 3D NOE-HMQC (refs. 35, 36). The advantages of the heteronuclear 3D HMQC experiments are the relative simplicity of the spectral analysis and the fact that they do not rely on the possibly weak J-coupling. However, they do require that the protein material is isotopically enriched in ^{13}C or ^{15}N which can be problematic. Here we will focus on the non-selective 3D NOE-HOHAHA experiment. In this experiment the increase in resolution is obtained by correlating resonances of ^1H nuclei - the most abundant nuclei in biomolecules - in three frequency domains. Furthermore the technique combines the measurement of J- and NOE-interactions in one experiment, and simultaneously registers a unique double magnetization transfer.

The 3D NOE-HOHAHA experiment can be described as the combination of the 2D NOE and 2D HOHAHA experiments, as shown in Fig. 4a. The FID in the t_3 time domain is recorded as a function of two variable times t_1 and t_2 , which are independently incremented. After Fourier transformation in the three dimensions a 3D NMR spectrum is obtained with three frequency axes. A cross peak in the 3D spectrum arises when magnetization of one proton is transferred in the first (NOE) mixing period to a second proton, and then in the second (HOHAHA) mixing period to a third proton.

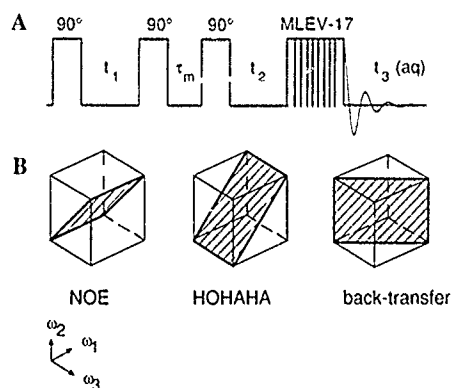


Fig. 4. A: Pulse sequence for a 3D NOE-HOHAHA experiment.

B: Cross-diagonal planes in a 3D NOE-HOHAHA spectrum.

Schematically the 3D spectrum can be presented in a cube with axes ω_1 , ω_2 and ω_3 . In such a 3D spectrum a body diagonal ($\omega_1=\omega_2=\omega_3$) can be identified, which

contains magnetization that was not transferred during any of the mixing periods. Additionally intensity will accumulate in three cross-diagonal planes, cf. Fig. 4b. In the 3D NOE-HOHAHA experiment the plane $\omega_2 = \omega_3$ (NOE plane) will contain magnetization transferred only during the first (NOE) mixing period. Similarly the plane $\omega_1 = \omega_2$ (HOHAHA plane) will correspond to the second coherence transfer step. Thus the 3D NOE-HOHAHA experiment contains the information present in the traditional 2D NOE and HOHAHA as a subset. Finally there is the unique $\omega_1 = \omega_3$ plane (back-transfer plane), which contains magnetization transferred during the NOE mixing period from proton *a* to *b* and then back to proton *a* during the HOHAHA mixing period.

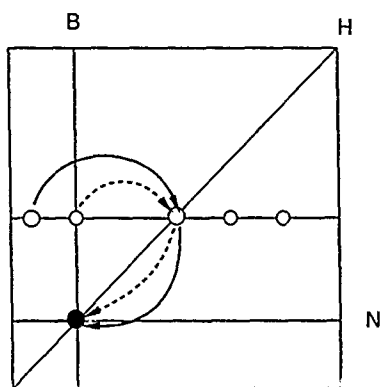


Fig. 5. The use of ω_3 cross sections at amide frequencies to obtain sequential assignments from 3D NOE-HOHAHA spectra. The lines indicated with N, H and B indicate the intersection with the NOE, HOHAHA and back-transfer planes, respectively. The solid arrows show the inter-residue $N\alpha N$ connectivity, the broken arrows the intra-residue $N\alpha N$ connectivity.

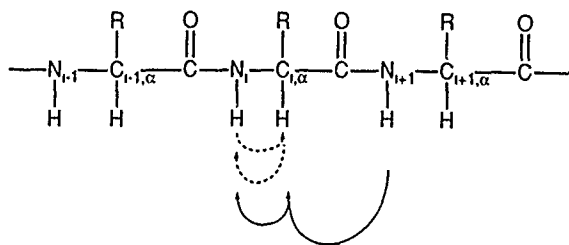


Fig. 6. Intra- and inter-residue connectivities along the backbone of a protein.

For the analysis of the 3D spectra we can take cross sections perpendicular to one of the axes, e.g. the ω_3 axis. The three diagonal planes will intersect this cross section at three lines indicated in Fig. 5 by N (NOE), H (HOHAHA) and B (back transfer). All cross peaks outside these three lines are due to a double magnetization transfer involving three different frequencies.

We will now explain how the information in such a NOE-HOHAHA spectrum can be used for the assignment of proton resonances in proteins, which is a prerequisite for NMR based structure determination. As outlined by Wüthrich (ref. 2) proton

resonances in proteins can be connected sequentially along the backbone chain by combining J-coupling and NOE information. Thus for each amino acid a series of intra-residue J-interactions and intra- and inter-residue NOE interactions can be identified, cf. Fig. 6. Fig. 5 shows schematically where the intra-residue and the sequential $C_{N\alpha N}[\text{NOE}, J]$ connectivities appear on a ω_3 cross section of a NOE-HOHAHA spectrum.

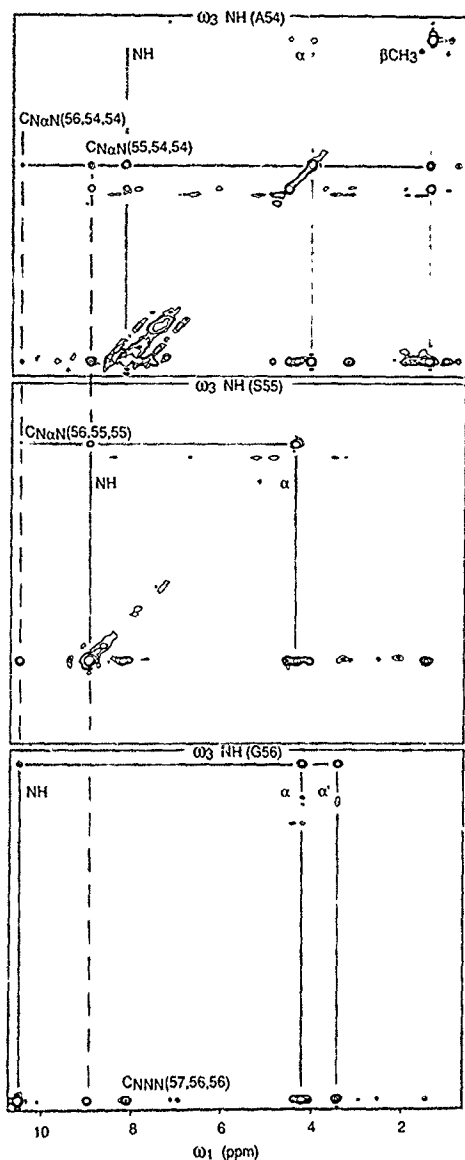


Fig. 7. ω_3 cross sections of the 3D NOE-HOHAHA spectrum of parvalbumin in H_2O at the NH frequencies of Ala-54 to Gly-56.

In fact many of such connectivities can be identified for the N, α and β protons along the backbone of a protein, the intensity of which depends on secondary conformation. An example of a sequential assignment is given in Fig. 7. It shows a

stretch of N α N connectivities in the NOE-HOHAHA spectrum of pike parvalbumin in H₂O, extending from Ala-54 to Gly-56.

It is clear that the combination of NOE transfer and HOHAHA mixing will be very useful for assignment procedures and identification of secondary structure elements in protein NMR spectra. Obviously the increased resolution in the 3D cube compared to the 2D plane is also of great help for identifying more NOE interactions, which can lead to a higher precision in structure determination or to solution of more complex structures. Recently we proposed the 3D NOE-NOE experiment as a method to study complex NOE networks (ref. 34). Apart from the fact that the method can be used to analyze spin diffusion networks and determine the cross-relaxation matrix with higher precision, the technique seems particularly useful for studying larger proteins. With higher molecular weights the HOHAHA mixing decreases in efficiency due to the short $T_{1\rho}$ relaxation time. The NOE transfer, on the contrary, becomes more efficient for larger proteins. The analysis of such spectra with many cross peaks is still complicated and further automation is highly desirable.

ACKNOWLEDGEMENTS

We like to thank Thea Koning for her work on the IRMA procedure and Jacob de Vlieg for his initial model building studies of the *lac* complex. We also gratefully acknowledge W.F. van Gunsteren and M. Billeter for providing their programs, J.H. van Boom for the synthesis of DNA fragments and M. Llinás for his contributions to the crambin studies.

REFERENCES

1. R.R. Ernst, G. Bodenhausen and A. Wokaun, Principles of Nuclear Magnetic Resonance in One and Two Dimensions, Clarendon Press, Oxford, 1987.
2. K. Wüthrich, NMR of Proteins and Nucleic Acids, Wiley, New York, 1986.
3. R. Kaptein, F. Boelens and J.A.C. Rullmann, Biomolecular structures from NMR: Computational aspects, in: W.F. van Gunsteren and P.K. Weiner (Eds.), Computer Simulation of Biomolecular Systems, Escom, Leiden, 1989, pp. 194-216.
4. R. Boelens, T.M.G. Koning and R. Kaptein, Determination of biomolecular structures from proton-proton NOEs using a relaxation matrix approach, J. Mol. Struct., 173 (1988) 299-311.
5. R. Boelens, R.M. Scheek, J.H. van Boom and R. Kaptein, Complex of *lac* repressor headpiece with a 14 base-pair *lac* operator fragment studied by two-dimensional nuclear magnetic resonance, J. Mol. Biol., 193 (1987) 213-216.
6. N. Lehming, J. Sartorius, S. Oehler, B. von Wilcken-Bergmann and B. Müller-Hill, Recognition helices of *lac* and λ repressor are oriented in opposite directions and recognize similar DNA sequences, Proc. Natl. Acad. Sci. USA, 85 (1988), 7947-7951.
7. I. Solomon, Relaxation processes in a system of two spins, Phys. Rev., 99 (1955) 559-565.
8. J.W. Keepers and T.L. James, A theoretical study of distance determinations from NMR. Two-dimensional nuclear Overhauser effect spectra, J. Magn. Reson., 57 (1984) 404-426.

9. M.S. Broido, T.L. James, G. Zon and J.W. Keepers, Investigation of the solution structure of a DNA octamer [d(GGAATTCC)]₂ using two-dimensional nuclear Overhauser enhancement spectroscopy, *Eur. J. Biochem.*, 150 (1985) 117-128.
10. E.T. Olejniczak, R.T. Gampe, Jr. and S.W. Fesik, Accounting for spin diffusion in the analysis of 2D NOE data, *J. Magn. Reson.*, 67 (1986) 28-41.
11. R. Boelens, T.M.G. Koning, G.A. van der Marel, J.H. van Boom and R. Kaptein, Iterative procedure for structure determination from proton-proton NOEs using a full relaxation matrix approach. Application to a DNA octamer, *J. Magn. Reson.*, 82 (1989) 290-308.
12. J.A.C. Rullmann, R.M.J.N. Lamerichs, C. Gonzalez, T.M.G. Koning, R. Boelens and R. Kaptein, Structure determination from NMR using a relaxation matrix approach; application to the solution structure of crambin, this volume.
13. T.M.G. Koning, R. Boelens and R. Kaptein, in preparation.
14. R.M.J.N. Lamerichs, R. Boelens, G.A. van der Marel, J.H. van Boom, R. Kaptein, F. Buck, B. Fera and H. Rüterjans, ¹H NMR study of a complex between the *lac* repressor headpiece and a 22 base pair symmetric *lac* operator, *Biochemistry*, 28 (1989) 2985-2991.
15. R. Kaptein and R. Boelens, NMR studies of repressor-operator interaction. The *lac* control region, in: F. Eckstein and D.M.J. Lilley (Eds.), *Nucleic Acids and Molecular Biology*, Vol. 2, Springer, Berlin, 1988, pp. 167-187, and references therein.
16. R. Kaptein, E.R.P. Zuiderweg, R.M. Scheek, R. Boelens and W.F. van Gunsteren, A protein structure from NMR data. *Lac* repressor headpiece, *J. Mol. Biol.*, 182 (1985) 179-182.
17. J. de Vlieg, R. Boelens, R.M. Scheek, R. Kaptein and W.F. van Gunsteren, Restrained Molecular Dynamics procedure for protein tertiary structure determination from NMR data: a *lac* repressor headpiece structure based on information on J-coupling and from presence and absence of NOEs, *Israel J. Chem.*, 27 (1986) 181-188.
18. C.O. Pabo and R.T. Sauer, Protein-DNA recognition, *Ann. Rev. Biochem.*, 53 (1984), 293-321.
19. R. Boelens, R.M. Scheek, R.M.J.N. Lamerichs, J. de Vlieg, J.H. van Boom and R. Kaptein, A two-dimensional NMR study of the complex of *lac* repressor headpiece with a 14 base pair *lac* operator fragment, in: W. Guschlbauer and W. Saenger (Eds.), *DNA-Ligand Interactions*, Plenum, New York, 1987, pp. 191-215.
20. R. Boelens, R.M.J.N. Lamerichs, J.A.C. Rullmann, J.H. van Boom and R. Kaptein, The interaction of *lac* repressor headpiece with its operator: an NMR view, *Prot. Seq. Data Anal.*, 1 (1988) 487-498.
21. S.R. Jordan and C.O. Pabo, Structure of the lambda complex at 2.5 Å resolution: details of the repressor-operator interactions, *Science*, 242 (1988) 893-898.
22. A.K. Aggarwal, D.W. Rodgers, M. Drott, M. Ptashne and S.C. Harrison, Recognition of a DNA operator by the repressor of phage 434: a view at high resolution, *Science*, 242 (1988) 899-907.
23. Z. Otwinowski, R.W. Schevitz, R.-G. Zhang, C.L. Lawson, A. Joachimiak, R.Q. Marmorstein, B.F. Luisi and P.B. Sigler, Crystal structure of *trp* repressor/operator complex at atomic resolution, *Nature*, 335 (1988) 321-329.
24. N. Lehming, J. Sartorius, M. Niemöller, G. Genenger, B. von Wilcken-Bergmann and B. Müller-Hill, The interaction of the recognition helix of *lac* repressor with *lac* operator, *EMBO J.*, 6 (1987) 3145-3153.
25. M. Billeter, T.F. Havel and I.D. Kuntz, A new approach to the problem of docking two molecules: the ellipsoid algorithm, *Biopolymers*, 26 (1987) 777.
26. J.A.C. Rullmann, R. Boelens and R. Kaptein, NMR based docking studies of *lac* repressor headpiece on a *lac* operator fragment, in: J. Gralla (Ed.), *DNA-Protein Interactions in Transcription*, UCLA Symposia on Molecular and Cellular Biology, new series, vol. 95, Alan R. Liss, New York, 1989, pp. 11-24.
27. J. de Vlieg, unpublished results.
28. W.F. van Gunsteren and H.J.C. Berendsen, *Groningen Molecular Simulation (GROMOS) Library Manual*, Biomos B.V., Nijenborgh 16, Groningen, 1987.
29. R.H. Ebright, Use of loss-of-contact substitutions to identify residues involved in an amino-acid - base pair contact: effect of substitution of Gln18 of *lac* repressor by Gly, Ser and Leu, *J. Biomol. Struct. Dyn.*, 3 (1985) 281.
30. J. de Vlieg, H.J.C. Berendsen and W.F. van Gunsteren, submitted for publication.

31. C. Griesinger, O.W. Sørensen and R.R. Ernst, Novel three-dimensional NMR techniques for studies of peptides and biological macromolecules, *J. Am. Chem. Soc.*, 109 (1987) 7227-7228.
32. H. Oschkinat, C. Griesinger, P.J. Kraulis, O.W. Sørensen, R.R. Ernst, A.M. Gronenborn and G.M. Clore, Three-dimensional NMR spectroscopy of a protein in solution, *Nature*, 332 (1988) 374-376.
33. G.W. Vuister, R. Boelens and R. Kaptein, Nonselective three-dimensional NMR spectroscopy. The 3D NOE-HOHAHA experiment, *J. Magn. Reson.*, 80 (1988) 176-185.
34. R. Boelens, G. W. Vuister, T.M.G. Koning and R. Kaptein, The observation of spin-diffusion in biomolecules by 3D NOE-NOE spectroscopy, *J. Am. Chem. Soc.* (1989) in press.
35. S.W. Fesik and E.R.P. Zuiderweg, Heteronuclear three-dimensional NMR spectroscopy. A strategy for the simplification of homonuclear two-dimensional NMR spectra, *J. Magn. Reson.*, 78 (1988) 588-593.
36. D. Marion, L.E. Kay, S.W. Sparks, D.A. Torchia and A. Bax, Three-dimensional heteronuclear NMR of ^{15}N -labeled proteins, *J. Am. Chem. Soc.*, 111 (1989) 1515-1517.
37. W. Gilbert and A. Maxam, cited in: J. H. Miller and W.S. Reznikoff (eds.), *The Operon*, Cold Spring Harbor Press, 1978, pp. 177-220.

DISCUSSION

POLTEV - What can you tell about distortion of DNA helix structure as a result of the interaction with the repressor ? Is it possible to characterize these distortions quantitatively by atom group displacements, by changes of base positions, torsion angles, etc. ?

RULLMANN - In most of our simulations we kept the DNA rigid in its energy minimized B-DNA form. We did this for two reasons : firstly, the calculations were done in vacuum with neutralized phosphates and residues. Structural predictions for DNA are then rather speculative. Secondly, at the time we did not have accurate NMR data on the operator, except that it has the same NOE pattern as B-DNA. Using IRMA we hope to be able to derive more quantitative information about the conformation adopted by the operator. Results obtained for free DNA fragments demonstrate that this is possible.

MILON - Have you ever used the IRMA technique in order to check the hypothesis of molecule rigidity, and more generally, could you comment about your approach of internal mobility.

RULLMANN - In the calculations on the DNA octamer we have seen that the largest constraint violations appeared on lower bounds involving sugar ring atoms. This correlates well with the assumption that the sugar rings are the most mobile parts of DNA. Preliminary results obtained by using mobility data taken from MD simulations to modify the correlation time locally, indeed show an improvement of the constraints. For crambin the mobility effects are (partly) obscured by the imprecision due to lack of stereospecific assignments. However, calculated NOEs tend to differ most from experimental values for residues, in the more flexible regions of the molecule. Local mobility effects are probably important here.

LAPLANTE - To what extent have 3D NMR techniques been practically useful for example for assignment purposes ?

RULLMANN - I must apologize for not being able to answer the question in detail, since I am not involved in the NMR work itself. At the moment the research is still very

much in the exploratory phase, working out possible assignment strategies e.g. for the NOE-NOE experiment. An important aspect is the possibility of automating the search for correlated cross-peak patterns and establishing possible assignments. Procedures for this are being developed in our group.

SOUMPASIS - Do you think that your methods are capable of detecting the absence or presence of H-bonding patterns in the DNA-protein recognition process ?

RULLMANN - Up to now we have only analyzed spectra of the complex obtained in D_2O , where the signals of protons in hydrogen bonds have disappeared because of chemical exchange with the solvent. In H_2O spectra - which are much more complex due to the larger number of resonances, and therefore have not been analyzed at present - hydrogen bond protons could in principle be observed provided exchange processes are slow enough. In practice most information on hydrogen bonds has to come from contacts between protons close to the hydrogen bond (such as observed between Gln-18 and G7) and from simulations.

RESTRAINED MOLECULAR MECHANICS OF ENKEPHALIN USING DISTANCES DERIVED FROM NMR (TRANSFERRED NOE)

A. MILON

Laboratoire de Chimie Organique des Substances Naturelles, associé au CNRS,
Centre de Neurochimie, 67084 Strasbourg, France.

SUMMARY

The membrane bound conformation of (D-Ala²)-Leu⁵ enkephalin has been studied by the transferred NOE technique in the presence of perdeuterated phospholipids. Seventy-two interproton distances have been determined and used as constraints in a molecular mechanics energy minimization using the TRIPOS force field. The refined model is in agreement with all the NMR data and is characterised by a type-II' β -turn around the last four residues and a γ -turn centered around D-Ala. The membrane bound conformation is closely related to the activity of the peptide.

INTRODUCTION

Leu⁵-enkephalin (Tyr-Gly-Gly-Phe-Leu) and Met⁵-enkephalin (Tyr-Gly-Gly-Phe-Met) are endogenous peptides with morphine like activity (ref. 1). Since their discovery, there have been extensive studies in order to determine their active conformation, because they are considered to bind to the same class of receptors as morphine, which has a rigid structure. Several studies in dimethyl sulfoxide solution have led to the conclusion that enkephalin existed in a β -turn structure involving the C-terminal four residues (refs. 2,3). Subsequent studies indicated that enkephalin was probably in equilibrium between folded and extended conformations (refs. 4,5). It was later shown that in aqueous solution the peptide is flexible and primarily takes extended conformations (refs. 4,6). Thus, the conformational analysis of peptides in solution gave little clue to the elucidation of conformation-activity relationship of enkephalins.

On the other hand, the importance of the interaction with phospholipid membranes is now well recognized. Thus for many peptides, the biological activities have been found to be related to the affinity to phospholipid membranes and the conformation in the membrane-bound state, rather than to the conformation in aqueous solution (ref. 7). In the particular case of enkephalins, anionic phospholipids are required for the binding to bilayers (ref. 8) but not for the binding to micelles (ref. 9). A bilayer membrane made of neutral and anionic phospholipids is also a more faithful model of cell membranes than are micelles.

However, the proton resonances of the membrane-bound peptide (total molecular weight higher than 10^7) are too broad to be analysed in detail by standard techniques. In the presence of phospholipid bilayers, the peptide exchanges rapidly between a membrane-bound state and a free state. The magnetization transfer is much more efficient in the membrane-bound state than in the free state and results in negative NOEs (because the peptide-vesicle system is in very slow motion). Therefore, the transferred NOE technique (TRNOE, ref. 10-12) allows the conformation determination of peptides in the membrane-bound state even though the spectrum of the bound peptide itself cannot be observed due to the slow motion.

We have succeeded recently in the elucidation of the specific conformation of [D-Ala²]-Leu⁵ enkephalin, a highly potent analog of enkephalin, in the membrane-bound state (ref. 13). We chose [D-Ala²]-Leu⁵ enkephalin because this peptide analog is 10 times more active than Leu⁵ enkephalin (ref. 14) and because the alanine methyl group was expected to reduce the conformational flexibility of the molecule and allowed a more clear-cut analysis of the TRNOE data (the presence of two glycine residues in Leu⁵-enkephalin makes it difficult to interpret TRNOE data unambiguously).

METHODS

In the presence of vesicles made of perdeuterated phospholipids (phosphatidylcholine-phosphatidylserine in a 1-1 molar ratio) in ¹H₂O solution, phase sensitive NOESY spectra have been observed at mixing times ranging from 50 to 300 ms. 1D NOE difference experiments were also performed using the 1-1 hard pulse technique for water suppression and a mixing time of 300 ms. Having performed the experiments in H₂O, complete sets of transferred nuclear Overhauser effect data have been obtained for pairs of N-H (all NH protons except the Tyr α -NH₃⁺ protons which exchange rapidly with H₂O protons), C- α (all α protons except Phe C- α proton whose resonance was masked by that of H₂O), C- β and side-chain C-H protons. These NOEs were converted into distances assuming that there was one rigid conformation (NOE proportional to d^{-6}) and using the tyrosine aromatic H(2)-H(3) distance as internal reference.

Using the SYBYL software (Tripos) implemented on a microVax II (DEC) and a graphic station PS390 (Evans & Sutherland), a starting conformation was built. Values close to the model proposed in ref. 13 were given to the dihedral angles. The distances derived from NMR were introduced as constraints of the type $E = k \cdot (d - d_0)^2$, where d is the interproton distance, d_0 is the distance used as a constraint and k is the force constant ($k = 200 \text{ cal. \AA}^{-2}$, a value comparable to the bond strength constant used in the Tripos force field). The energy of the constrained molecule was minimized using the Tripos force field maximin 2 and

compared to the minimized energy of the starting conformation. The final conformation energy was minimized again after removal of the constraints in order to assess the strain existing in the final molecule.

RESULTS AND DISCUSSION

After conversion of the NOE data into distances, a set of 17 interproton distances and of 53 lower limits of interproton distances was available for restrained molecular mechanics energy minimization (table 1).

TABLE 1

Distances used as constraint in the energy minimization (in Å)^a

d:NHAla-	NHGly=3.0 αTyr>3.6 βLeu>4	NHLeu >4 αAla=2.6 MetAla=3.1	NHPhe>4 αGly>4	ArTyr>4 βTyr=3.5	αLeu>4 βPhe>4
d:NHGly-	NHPhe=3.3 αAla=3.1 MetLeu>4	ArPhe>4 αGly=2.7 MetAla=3.1	ArTyr>4 βTyr>4	αLeu>4 βPhe>4	αTyr>4 βLeu>4
d:NHLeu-	NHPhe=2.6 αAla>4	ArPhe=3.9 αGly>4	ArTyr>4 βTyr>4	αLeu=2.8 βPhe>4	αTyr>4 MetAla>4
d:NHPhe-	ArPhe=3.7 αGly=2.6	ArTyr>4 βTyr>4	αLeu>4 βPhe=3.0	αTyr>4 βγLeu=3.5	αAla>4 MetAla>4
d:ArPhe-	ArTyr>4 βTyr>4	αLeu>4	αTyr>4	αAla>4	αGly>4
d:ArTyr-	αLeu>4	αTyr>4	αAla>4	αGly>4	βPhe>4
d:αLeu-	αTyr>4 MetAla>4	αAla>4	αGly>4	βTyr>4	βPhe>4
d:αTyr-	αAla>4	αGly>4	βPhe>4		
d:αAla-	αGly>4	βPhe>4			
d:αGly-	βTyr>4	βPhe=3.7			
d:βTyr-	MetAla>4				
d:βγLeu-	3.7<MetAla<4				

^a When several protons were involved, as for α Gly for instance, the closest proton was taken

From the NOE data, a model was built "by hand", using molecular models. It was used as a starting conformation for the energy minimization process. Random conformations, and structures derived from X-Rays, were also tested. They generally gave either problems during the energy minimization (rising energies) and (or) very distorted structures having unrealistic bond angles and bond distances and very high energies. In this approach, the validity of the model was

demonstrated by the fact that the energy of the constrained molecule was close to the energy of the structure minimized in the absence of constraint. After energy minimization the constrained distances were equal to those indicated in table 1 within 0.1 Å.

From the refined structure a coordinate file was created (.mol file, which is available upon request) and used by the Moldraw software (J.M. Cense) on a Macintosh computer to generate the ball and stick model shown in figure 1.

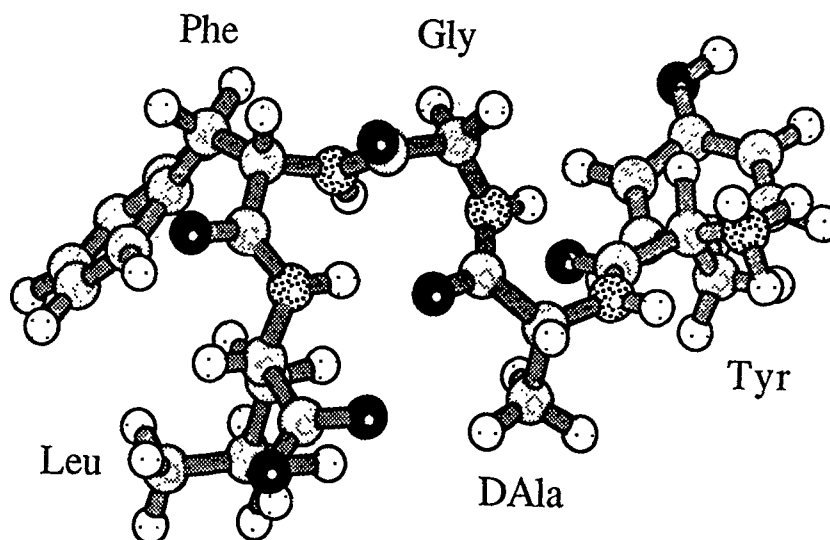


Fig. 1. Membrane bound conformation of (D-Ala²)-Leu⁵ enkephalin as determined by the combined use of transferred NOE and restrained energy minimization.

The refined model obtained after energy minimization was found to be close to what was guessed on the basis of molecular models and the main conclusions were confirmed. In particular the type II' β -turn between D-Ala, Gly, Phe, Leu, the γ -turn around D-Ala and the position of the tyrosine apart from the rest of the molecule were preserved. The membrane-bound conformation of enkephalins approximates very closely that of highly active cyclic analogs (ref. 15), and appears to be correlated with activity (ref.13).

The use of molecular mechanics has confirmed that our model is consistent with all the NOE data and it is therefore reasonable to believe that there is one major membrane bound conformation of enkephalins.

ACKNOWLEDGEMENTS

The work on which this work is based was performed in professor T. Miyazawa's laboratory (University of Tokyo, Japan, ref. 13).

The author thanks Professor C.G. Wermuth for the use of molecular modelling facilities and professors G. Ourisson and Y. Nakatani for their encouragements.

REFERENCES

- 1 J. Hughes, T.W. Smith, H.W. Kosterlitz, L.A. Fothergill, B.A. Morgan, and H.R. Morris, Identification of two related pentapeptides from the brain with potent opiate agonist activity, *Nature*, 258 (1975) 577-579.
- 2 C.R. Jones, W.A. Gibbons, and V. Garsky, Conformation of Met⁵-enkephalin determined by high field PMR spectroscopy, *Nature*, 262 (1976) 778-782.
- 3 E.R. Stimson, Y.C. Meinwald, and H.A. Scheraga, Solution conformation of enkephalin. A Nuclear Magnetic Resonance study of ¹³C-enriched carbonyl carbons in [Leu⁵]-enkephalin, *Biochemistry*, 18 (1979) 1661-1671.
- 4 T. Higashijima, J. Kobayashi, U. Nagai, and T. Miyazawa, Nuclear-magnetic-resonance study on Met-enkephalin and Met-enkephalinamide, *Eur. J. Biochem.*, 97 (1979) 43-57.
- 5 D. Marion, C. Garbay-Jaureguiberry, and B.P. Roques, Correlation of ¹⁵N and ¹³C relaxation data as a mobility probe for linear peptides. A theoretical and experimental study using enkephalins and related models, *J. Magn. Reson.*, 53 (1983) 199-212.
- 6 M.A. Khaled, D.W. Urry, and R.J. Bradley, *J. Chem. Soc. Perkin Trans., II* (1979) 1693-1699.
- 7 K. Wakamatsu, T. Higashijima, and T. Miyazawa, Active conformations of membrane-bound peptides, In: H. Imura, T. Goto, T. Nakajima and T. Murachi (Eds.), *Natural Products and Biological Activities*, Elsevier, Amsterdam, 1986, pp. 237-244.
- 8 H.C. Jarrell, R. Deslaurier, W.H. McGregor, and I.C.P. Smith, Interaction of opioid peptides with model membranes. A carbon-13 nuclear magnetic study of enkephalin binding to phosphatidylserine, *Biochemistry*, 19 (1980) 385-390.
- 9 B.A. Benham, and C.M. Deber, Evidence for a folded conformation of methionine- and leucine-enkephalin in a membrane environment, *J. Biol. Chem.*, 259 (1984) 14935-14940.
- 10 K. Wakamatsu, A. Okada, T. Higashijima, and T. Miyazawa, NMR analysis of the conformations of membrane-bound peptides, *Biopolymers*, 25 (1986) S193-200.
- 11 G.M. Clore, and A.M. Gronenborn, Theory of the time dependent transferred nuclear Overhauser effect : applications to structural analysis of ligand-protein complexes in solution, *J. Magn. Reson.*, 53 (1983) 423-442.
- 12 K. Wakamatsu, A. Okada, T. Miyazawa, Y. Masui, S. Sakakibara, and T. Higashijima, Conformations of yeast α -mating factor and analog peptides as bound to phospholipid bilayer. Correlation of membrane-bound conformation with physiological activity, *Eur. J. Biochem.*, 163 (1987) 331-338.
- 13 A. Milon, T. Miyazawa and T. Higashijima, Transferred NOE analyses of membrane-bound enkephalin analogs by ¹H-Nuclear magnetic resonance :

correlation between activities and membrane-bound conformations, Biochemistry, in press.

- 14 J.S. Morley, Structure-activity relationships of enkephalin-like peptides, Ann. Rev. Pharmacol. Toxicol., 20 (1980) 81-110.
- 15 H.I. Mosberg, R. Hurst, V.J. Hruby, K. Gee, H.I. Yamamura, J.J. Galligan, and T.F. Burks, Bis-penicillamine enkephalins possess highly improved specificity towards δ opioid receptors, Proc. Natl. Acad. Sci. USA., 80 (1983) 5871-5874.

DISCUSSION

FAUPEL - Conformation membranaire de Leu-Enkephaline : how is it incorporated in the membrane, i.e. you have found the conformation, but how is it oriented in the membrane ?

MILON - As we are using perdeuterated phospholipids, we get no information at the moment about the position in the membrane. Only the membrane bound conformation of the peptide is determined. We are currently developing new methodologies to answer this very important question.

PRELIMINARY STRUCTURE DETERMINATION OF TWO TOXINS USING NMR DATA

A. MIKOU¹, S. LAPLANTE¹, R. LE GOAS², M.-A. DELSUC², I. CHARPENTIER³,
E. GUITTET^{1*} AND J.-Y. LALLEMAND².

¹Laboratoire de RMN, ICSN, CNRS, 91190 Gif sur Yvette (France).

²Laboratoire de RMN, DCSO, Ecole Polytechnique, 91128 Palaiseau Cedex (France).

³Service de Biochimie, CEN, 91191 Saclay, Gif sur Yvette (France).

SUMMARY

A rapid means of determining a preliminary three-dimensional structure of a cobra toxin (71 aa) using NMR data is described. First, secondary structures were easily assigned using characteristic MCD patterns. Distance constraints from these secondary structures were then applied in restrained energy minimization and molecular dynamic calculations to provide a preliminary three-dimensional reconstruction of the cobra toxin. A similar study is currently underway for a scorpion toxin (64 aa).

Abbreviations : NMR, Nuclear Magnetic Resonance; MCD, Main Chain Directed; HOHAHA, total correlation spectroscopy; COSY, Correlated Spectroscopy; NOESY, Nuclear Overhauser Effect Spectroscopy.

INTRODUCTION

NMR has become an important tool in determining three-dimensional structures of small proteins in solution. The key factor involved is the nuclear Overhauser effect which reveals interproton short and long range distances of less than 5 Å. To determine possible conformations which satisfy all these distance constraints, restrained energy minimization and molecular dynamic calculations need to be used (refs. 1-2). However, many difficulties quickly arise. One of the most serious is to assign the NMR crosspeaks in the 2D NOE maps to the parent protons.

To solve this assignment problem, several methods have been proposed. The sequential assignment method (ref. 3) consists first of identifying all or most amino acid residues using HOHAHA and COSY experiments which reveal the connectivities through bonds. Using α H-NH NOE information, the neighboring residues can then be identified. The resulting short segments of neighboring residues are then matched against the corresponding segments in the primary sequence. Secondary structures are thus identified only after the amino acid assignment has been completed.

The recently proposed main-chain directed (MCD) assignment strategy (refs. 4-5), on the other hand, requires defining all the main-chain spin systems followed by recognition of NOE connectivity patterns which are characteristic of secondary structures (helices, sheets, turns and extended chain). The recognition of a few amino acid types places the discovered secondary structure within the polypeptide sequence. Unlike the

sequential assignment approach, the MCD assignment strategy does not rely on the difficult task of recognizing many side-chain systems among the crowded side-chain regions of the NMR maps.

In this paper, we have used MCD patterns to characterize secondary structures of both the cobra and the scorpion toxins. Finally, in the case of the cobra toxin, distance constraints from the discovered secondary structures were applied in restrained calculations to determine a preliminary three-dimensional structure.

METHODS

Samples preparation

The solutions of the cobra toxin (α -cobratoxin of the long chain toxins of *Naja Naja Siamensis*) and the scorpion toxin (AaH III toxin from *Androctonus Australis* Hector) were unbuffered and contained in 5-mm NMR sample tubes. The protein concentrations were 3 mM and 2.5 mM for the cobra and the scorpion toxins, respectively.

Nuclear Magnetic Resonance

All the ^1H NMR experiments were performed at 400 and 600 MHz on Bruker AM spectrometers. The data obtained were transferred to a Vax GPX II station and subsequently processed with the GIFA software (ref. 6).

Energy Calculations

All the calculations were performed on an IBM 3090 VF system using the AMBER software (refs. 7-8).

DISCUSSION

The MCD patterns prove to be particularly useful to characterize secondary structures such as β sheets and helices.

The β sheet of the cobra toxin was characterized by first observing the readily apparent αH - αH crosspeak shown as peak number 1 in the NOESY spectrum in Fig. 1. The αH - αH peaks are usually more visible in spectra taken in D_2O . Peak number 1 was then used as the origin for locating the inner-loop MCD pattern shown as thick lines (and numbered). The numbers coincide with the numbered peaks in the inset. Aligned along both αH chemical shifts of peak 1, four crosspeaks can be found which form a rectangular shape (see peaks 2, 3, 4, and 5). This rectangle has alternating strong (peaks 2 and 4) and weak (peaks 3 and 5) crosspeak intensities. For this analysis, only the NH - αH part of the spin system needed to be defined in order to distinguish inter-residue from intra-residue crosspeaks. Connected inner-loops and loop subsets (see (ref. 4) for further discussion on loop subsets) were subsequently found thus extensively defining a triple-stranded antiparallel sheet. Alignment of this complete pattern with the primary sequence for assignment purposes was straightforward since only a few easily

identifiable side-chain spin systems needed to be used (i.e. Val, Thr, Gly etc).

MCD patterns can also be used to identify helix secondary structure. The basic MCD patterns of the α helix consists of a series of closed loops, a single loop is formed by a strong NH-NH NOE connectivity and also intra-residue and inter-residue NH- β H connectivities (ref. 4). Connecting five of these basic loops, we then characterized a mini-helix in the cobra toxin. The helix was further confirmed by the presence of α H-NH ($i, i+3$) and α H- β H ($i, i+3$) NOE connectivities (ref.3). The amino acids within the helix were assigned by aligning the complete pattern with the primary sequence.

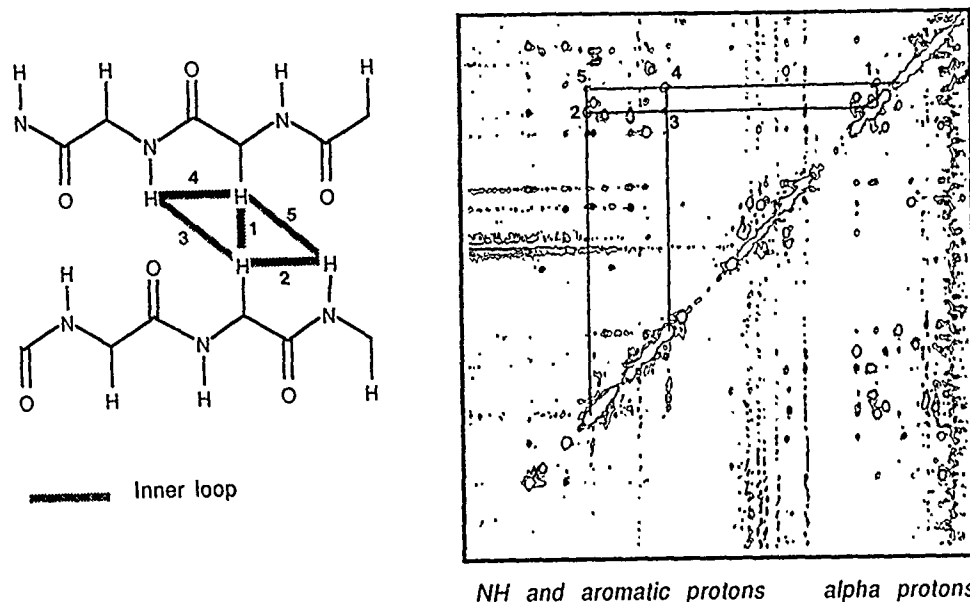


Fig. 1. 600 MHz NOESY subspectrum (300 msec. mixing time) of cobra toxin in D_2O . The number beside the crosspeaks correspond to the numbered striped lines of the full-loop MCD pattern in the inset.

The MCD based assignments represent 21 residues involved in the triple-stranded antiparallel β sheet (strand 1 : from residue 52 to 58; strand 2 : 19 to 25; strand 3 : 36 to 42) and 7 residues involved in the mini-helix (from residue 29 to 34), that is altogether 28 residues of the 71 amino acids representing more than one third of the cobra toxin sequence.

The same triple stranded antiparallel β sheet is readily observed in the XRAY structure, however the mini-helix structure which is characterized by the MCD patterns is not particularly mentioned in the crystal structure (ref.9).

The scorpion toxin was subjected to a similar analysis as described; a single triple-stranded antiparallel β sheet and an α helix of two and a half turns were found. Although the crystal structure has not been determined, toxins of the same family also exhibit a β sheet and an α helix in similar regions of the primary sequence, as revealed

by NMR (ref. 10) and XRAY studies (refs. 11-13).

In the case of the cobra toxin, a set of 70 NMR distance constraints collected from the secondary structures, was used as input in energy minimization and molecular dynamic calculations, using the AMBER program (refs. 7-8). This study allowed us to determine a preliminary three dimensional structure of the cobra toxin.

The XRAY structure was used as a starting point for all these calculations. In the first calculation, the XRAY structure was energy minimized using a harmonical potential.

$$V(d) = 0.5 k_d (d - d_{mes})^2$$

where d_{mes} is an estimated distance based on the observed NOE between the parent protons and d is the target distance (refs. 14-15). For a strong, a medium, a small NOE, d_{mes} was taken as 2.4Å, 2.9Å and 3.75Å, respectively. k_d is taken as 25.0, 17.5 and 11.0 kcal/moleÅ for the strong, medium and small NOEs. For distances involving methylene protons (for which stereospecific assignments remain presently to be made), k_d was taken as 1.6 kcal/moleÅ.

For the restrained molecular dynamics, the following three steps were performed :

- the distance restraints were first introduced and the starting structure again energy minimized.

- a first run of molecular dynamics (1.25ps at 300K, consisting of 0.5fs steps) equilibrated the system thermally.

- a second run of molecular dynamics (5ps at 300K, consisting of 0.5fs steps). The average structure is then computed and minimized and analyzed.

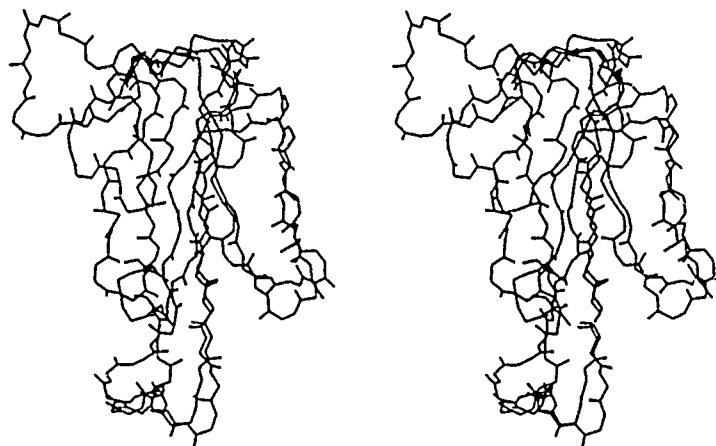


Fig. 2. Superposition of the restrained energy minimized mean structure with the initial X-ray structure (backbone) of α cobratoxin.

Among the salient features of the reconstructed structures, as compared with the starting X-ray structure, one should note first the striking similarity of the two structures

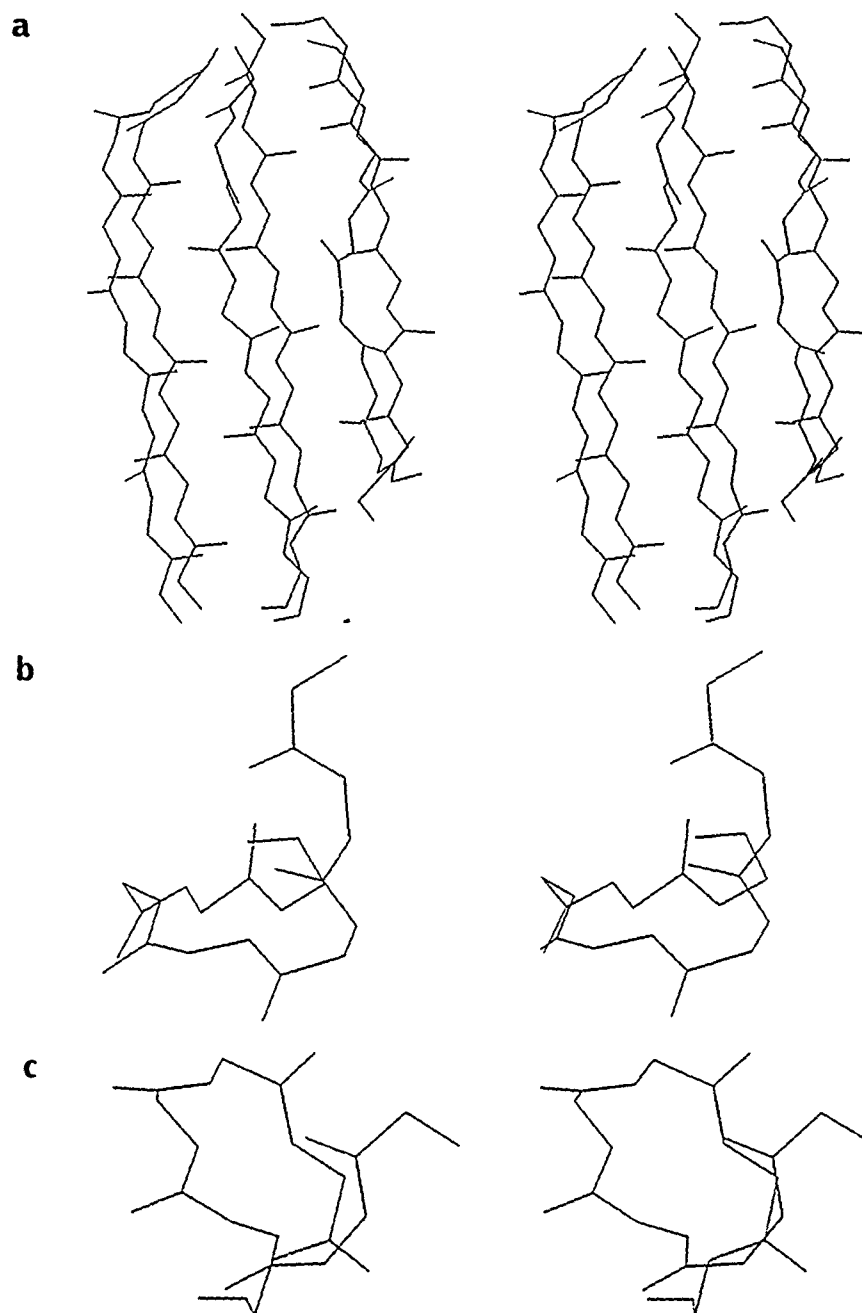


Fig. 3. Details of the energy minimized X-ray and final structures of α cobratoxin. (a) The superimposed antiparallel β sheets. (b) α helix (energy minimized X-ray structure). (c) α helix (final structure).

with respect to the overall shape, size, polypeptide fold and secondary structure (Figure 3). To go into some details, one could mention in addition :

- a good overall agreement between the X-ray and the calculated structure at the triple-stranded antiparallel β sheet, although the strand extending from residue 36 to residue 40 is more affected (Figure 4).

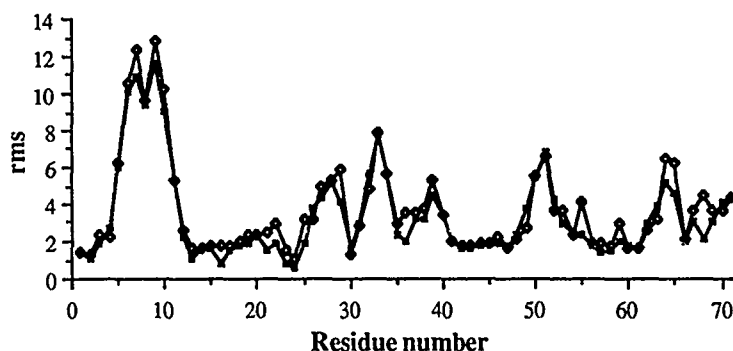


Fig. 4. Atomic rms difference between the restrained energy minimized mean structure with the initial X-ray structure of a cobra toxin. Diamonds represent the average RMS over all the atoms, whereas squares represent the corresponding RMS over the backbone.

- the presence of a rather canonical mini-helix between residue 29 and 35 which is hardly recognized in the crystal structure (Figure 4).

- the marked differences between the two structures, significantly enough, occur in regions where intermolecular interactions are expected in the crystal (Phe-29 of one molecule is reported to "protrude and tuck into the arch formed by the tail residues (63-68) and the underside of the first loop (residues 3-10) of the neighboring molecule") (ref. 9).

CONCLUSION

The present results allow only preliminary and careful conclusions to be drawn due to the incomplete analysis of the NMR data. They clearly demonstrate however that a rough analysis of the NMR data set, merely based on the MCD approach, allows to give a first insight into the three-dimensional structure of small sized proteins and a first refinement of the available X-ray structures when the resolution is relatively poor.

ACKNOWLEDGMENTS

Financial support was provided by the Chateaubriand Scholarship (French government) to S.R.L. and by Roussel Uclaf to A.M.

We thank Dr.Menez for providing the cobra toxin and Roussel Uclaf for providing the scorpion toxin, Bruker for providing us 600 MHz spectrometer time.

The extensive calculations were made possible by the collaboration of the CNRS and IBM within the "Groupement Scientifique de Modélisation Moléculaire".

REFERENCES

- 1 R.Kaptein, E.R.P.Zuiderweig, R.M.Sheek, R.Boelens and W.van Gunsteren, J. Mol. Biol. 182 (1985) 179-182.
- 2 G.M.Clôre, A.M.Gronenborn, A.T.Brunger and M.Karplus, J. Mol. Biol. 186 (1985) 435-455.
- 3 K.Wüthrich, (1986) NMR of Proteins and Nucliec Acids, Wiley, New York.
- 4 S.W.Englander and A.J.Wand, Biochemistry 26 (1987) 5953-5958.
- 5 D.L.Di.Stephano and A.J.Wand, Biochemistry 26 (1987) 7272-7281.
- 6 Available from the authors upon request.
- 7 S.J.Weiner, P.A.Kollman, D.A.Case, U.C.Singh, C.Ghio, G.Alagona, S.Profeta and P.Weiner, J. Amer. Chem. Soc. 106 (1984) 765-784.
- 8 S.J.Weiner, P.A.Kollman, D.A.Case and D.T.Nguyen, J. Comp. Chem. 7(2) (1986) 230-252.
- 9 M.D.Walkinshaw, W.Saenger and A.Maelicke, Proc. Natl. Acad. Sci. USA 77 (1980) 2400-2404.
- 10 D.G.Nettesheim, R.Klevit, G.Drobny, D.D.Watt and N.R.Krishna, Biochemistry 28 (1989) 1548-1555.
- 11 Almassy, R.J., Fontecilla-Camps, J.C., Suddath, F.L. and Bugg, C.E. (1983) J. Mol. Biol. 170, 497.
- 12 J.C.Fontecilla-Camps, R.J.Almassy, S.E.Ealick, F.L.Suddath, D.D.Watt, R.J.Feldman and C.E.Bugg, Trends Biochem. Sci. 6, (1981) 291.
- 13 J.C.Fontecilla-Camps, R.J.Almassy, F.L.Suddath and C.E.Bugg, Toxicon (1982) 20.
- 14 M.P.Williamson, T.F.Havel and K.Wütrich, J. Mol. Biol. 182 (1985) 295-315.
- 15 Havel, T.F., Crippen, G.M. and Kuntz, I.D. (1979) Biopolymers 18, 73-81.

DISCUSSION

SURCOUF - How many NOE data did you use in your dynamics calculations ?

MIKOU - In the case of the α Cobra-toxin, the distance constraints were collected from the antiparallel β sheet and the small helix. About 70 interproton distances were used in the restrained calculations.

MORIZE - What is the resolution of the X-ray structure ? and why are the RMS values so big ? (about 6 Å for an helix between the model structure and the X-ray structure) ?

MIKOU - Unfortunately the X-ray structure has a low resolution about 2.8 Å. Given this low resolution, we were interested in refining the structure using NMR data. This is why, we used the X-ray structure as a starting point in restrained molecular dynamic calculations.

Now concerning the graph which represents the RMS between the X-ray structure and the model structure (X-ray structure + restrained molecular dynamic calculations using the NMR distance constraints), we can see that the anti-parallel β sheet of the α cobra-toxin remains relatively stable during the calculations.

However, the higher RMS values indicate that the large differences occur for the region between residue 29 to 35.

This region corresponds to mini-helix which was characterized by MCD patterns but it was not mentioned in the publication of the X-ray structure (M.D. Walkinshaw et al, Proc. Natl. Acad. Sci. USA 77 (1980) 2400-2404).

Finally NMR distance constraints collected from this region, were applied in the restrained molecular dynamics, we then obtained a small helix, whereas in the crystal structure it is not as well formed.

PEPE - Several structures of snake venom toxins were solved by X-ray analysis, they are very similar and they only differ by the length of the large loop. These structures do not present any α -loop as indicated by NMR analysis performed on one of these toxins, let me notice that α -loop are very well solved by X-ray analysis.

MIKOU - As I mentioned, the distance constraints from the helix region were applied in the restrained calculations. We found that the resulting local structure which satisfy all these conformational restraints was a mini-helix. This helix is not mentioned in X-Ray structure probably because of the low resolution (2.8 Å). Subsequently, in the X-Ray structure intermolecular interactions were expected in this region, this probably destabilize the small helix.

WODAK (comment) - A detailed analysis of contacts between molecules in the cobra toxin crystal structure could be useful in explaining the difference between the solution and crystal structures of this protein. It may well be possible that the region shown to adopt a helical conformation in solution may be prevented from doing so in the crystal due to interactions with neighbouring molecules, a fact which one tends to forget.

**BIS-INTERCALATION OF DITERCALINIUM AND "FLEXIBLE"
ANALOGUE IN THE OCTANUCLEOTIDE d(TpTpCpGpCpGpApA)₂:
A COMPARATIVE STUDY BY NMR AND MOLECULAR MODELLING**

Joël POTHIER¹, Muriel DELERIERRE², Marie-Christine BARS², Christiane
GARBAY², Jean IGOLEN³, Marc LE BRET¹ & Bernard P. ROQUES².

¹ Laboratoire de Physico-Chimie Macromoléculaire, URA 158 CNRS, Institut
Gustave Roussy, rue Camille Desmoulins, 94800 VILLEJUIF.

² Département de chimie Organique U266 INSERM, UA 498 CNRS, UER des
Sciences Pharmaceutiques et Biologiques, 4 Avenue de l'Observatoire, 75006 PA-
RIS France.

³ Unité de chimie Organique, Institut Pasteur, 28 rue du Docteur ROUX 75015
PARIS France

SUMMARY

The structures of the complexes formed between both ditercalinium (7H-pyrido[4,3c]carbazole rigid dimer) and a ditercalinium analogue with a flexible chain, with the self-complementary octanucleotide d(TpTpCpGpCpGpApA)₂ have been investigated by 400 MHz ¹H and 160 MHz ³¹P NMR. The assignation of nearly all protons and all phosphorus atoms was achieved by 2D NMR, and NOE intermolecular effects show unambiguously bis-intercalation of both drugs into the octanucleotide.

INTRODUCTION

7H-Pyrido [4,3c] carbazole dimers of the ditercalinium family are DNA bis-intercalators that display high DNA affinity and antitumor properties. (ref 1) Their antitumor properties depend on substitution on the pyridocarbazole rings and/or "flexibility" of the linking chain. (ref 2) When the two piperidine rings of ditercalinium (figure 1 a) are replaced by six methylene groups (as shown in figure 1 b), the drug's cytotoxicity for the L1210 cells disappears and it no longer displays antitumour properties in mice. However, the K affinity for DNA is 2.10^7 M^{-1} for ditercalinium and 1.10^8 M^{-1} for the flexible analogue.

To get a better understanding in the inverse order in affinities and cytotoxicities, the interaction of ditercalinium and its "flexible" analogue with both the tetranucleotide d(CpGpCpG)₂ and the octanucleotide d(TpTpCpGpCpGpApA)₂ has been investiga-

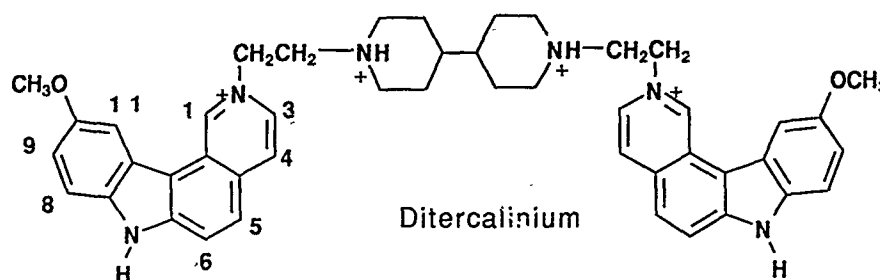


Figure 1 a

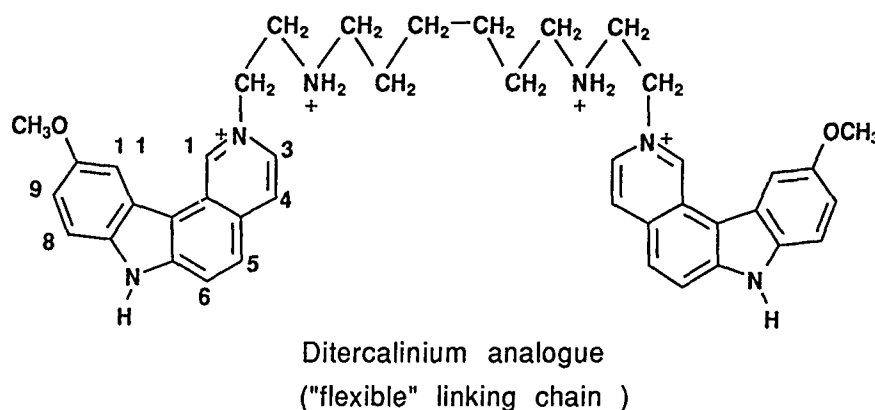


Figure 1 b

MATERIALS AND METHODS

MATERIALS

The oligonucleotides samples were chromatographed over chelex 100 resin to remove dicationic metal ions. The samples (nucleotides and drugs) were then lyophilized twice from D₂O prior to use and redissolved in 0.05 M deuterioacetate buffer (from CEA deuterioacetic acid) pH 5.8 to a 1 mM final concentration in duplex. Concentrated drug solution was gradually added to the oligonucleotide solutions to a 1:2 drug to helix ratio for identification of ¹H and ³¹P bound signals from 2D exchange spectroscopy and then to a 1:1 drug to helix ratio for structural studies. All experiments were carried out at 295°K, except for 2D ³¹P exchange experiments which were recorded at various temperatures from 295°K to 315°K.

^1H NMR

Proton NMR spectra were obtained using a 400 MHz Bruker AM spectrometer equipped with an ASPECT 3000 computer. Mixing times from 50 ms to 300 ms were used in 2D NOESY experiments.

^{31}P NMR

Phosphorus spectra were obtained with a selective 5 mm phosphorus-proton dual probe, at 162 MHz, under the following experimental conditions: 1D spectra were recorded with broad-band proton decoupling using a composite pulse decoupling system applied during acquisition, the two dimensional $^{31}\text{P}\{^1\text{H}\}$ chemical shift correlated experiments were performed using the polarization transfer from proton to phosphorus via the phosphorus proton coupling constant or using the reverse experiment of transfer from phosphorus to proton. The two dimensional ^{31}P - ^{31}P chemical exchange experiments were improved using a composite pulse decoupling during the acquisition and mixing times of 300 and 500 ms were used.

RESULTS

^1H Resonances assignments

In all cases, one of the major features of the ^1H spectra was that symmetry of the self complementary oligonucleotides was not perturbed by drug fixation (as seen previously for ditercalinium in $d(\text{CGCG})_2$, (ref 3)), i.e. only one signal could be seen for the two symmetrical protons on the two nucleotide strains, and the dimer drug protons remain symmetric in complexes.

The 2D COSY experiments led to the assignment of all spin systems in free and bound nucleotides and the 2D NOESY experiments permitted the nucleotides to be ordered through sequential assignment. In the case of $d(\text{CpGpCpG})_2$, only the free form of tetranucleotide and the complex with flexible analogue was investigated, the complex with ditercalinium having been previously studied by Delbarre et al. (Ref 3). Bound tetranucleotide protons were assigned by exchange peaks in 2D NOESY of the 1:2 drug to helix ratio solution, which correlated free (previously assigned) and bound protons. With the octanucleotide complex, no exchange peaks were observed under the same conditions (Temperature: 295°K, mixing time: 300ms), so identification of spin systems and sequential assignment were carried out using the 1:1 drug to helix ratio solution. For the tetranucleotide complex, a 1:1 drug to helix ratio could not be reached because of precipitation of the complexes.

The proton chemical shifts in all complexes (Tables 1 and 2) show that the aromatic protons of both nucleotide and drug are nearly equally shifted for ditercalinium complexes and analogue complexes, except for drug proton H5 (and H8 and H9 in the octanucleotide complexes). In contrast, nucleotide ribose protons differed much more between ditercalinium complexes and analogue complexes. For $d(\text{CGCG})_2$ complexes, the largest chemical shift differences observed between ditercalinium complex and analogue complex, was those of the internal cytosine and guanine ri-

TABLE 1

Chemical shifts (δ) in octanucleotide free and in complexes

ribose protons	free	analog complex	diter- calinium complex	$\Delta\delta^*$	aromatic protons	free	analog complex	diter- calinium complex	$\Delta\delta^*$
1Thy1'	6.05	5.87	5.87	0.00	1ThyCH3	1.78	1.65	1.64	0.01
1Thy2'	2.20	2.23	2.20	0.03	1Thy6	7.61	7.60	7.57	0.03
1Thy2"	2.57	2.45	2.45	0.00					
2Thy1'	6.17	5.93	5.93	0.00	2ThyCH3	1.73	1.62	1.60	0.02
2Thy2'	2.28	2.22	2.16	0.06	2Thy6	7.62	7.59	7.57	0.02
2Thy2"	2.58	2.43	2.42	0.01					
3Cyt1'	5.61	5.83	5.62	0.21	3Cyt5	5.73	5.55	5.51	0.04
3Cyt2'	2.13	2.27	2.29	0.02	3Cyt6	7.57	7.70	7.65	0.05
3Cyt2"	2.42	2.70	2.92	0.22					
4Gua1'	5.88	5.33	5.38	0.05	4Gua8	7.92	7.64	7.65	0.01
4Gua2'	2.69	2.49	2.31	0.18					
4Gua2"	2.63	2.49	2.31	0.18					
5Cyt1'	5.58	5.95	5.92	0.02	5Cyt5	5.34	5.09	5.07	0.02
5Cyt2'	1.77	2.27	2.27	0.00	5Cyt6	7.24	7.27	7.26	0.01
5Cyt2"	2.21	2.54	2.54	0.00					
6Gua1'	5.33	5.22	5.13	0.09	6Gua8	7.83	7.64	7.65	0.01
6Gua2'	2.56	2.52	2.50	0.02					
6Gua2"	2.60	2.52	2.50	0.02					
7Ade1'	5.98	5.88	5.88	0.01	7Ade8	8.06	7.96	7.95	0.00
7Ade2'	2.56	2.47	2.46	0.01					
7Ade2"	2.76	2.69	2.92	0.24					
8Ade1'	6.19	6.16	6.13	0.02	8Ade8	8.07	8.01	8.01	0.00
8Ade2'	2.52	2.51	2.50	0.01					
8Ade2"	2.33	2.33	2.32	0.02					

* $\Delta\delta = |\delta(\text{analog}) - \delta(\text{ditercalinium})|$

TABLE 2

Chemical shifts (δ) of drug aromatic protons in complexes

	analog	ditercalinium]	$\Delta\delta^*$		analog	ditercalinium]	$\Delta\delta^*$
H1	9.50	9.55	0.05	H8	6.50	6.59	0.09
H3	8.21	8.24	0.03	H9	6.22	6.12	0.10
H4	7.67	7.59	0.08	H11	7.31	7.28	0.03
H5	7.51	7.29	0.22	H6	8.00	7.92	0.08
CH3	3.69	3.63	0.06				

* $\Delta\delta = |\delta(\text{analog}) - \delta(\text{ditercalinium})|$

bose protons. For $d(\text{TTCGCGAA})_2$ complexes, the largest differences occur for the the cytosine 3 and guanine 4 ribose protons. Interestingly, these latter protons are on the same side of intercalation site as the drug proton H5 which is also differently shifted in the two complexes (see above). A similar geometry at the intercalation sites for ditercalinium and analogue complexes is suggested by the similarity of the chemical shifts of the aromatic protons

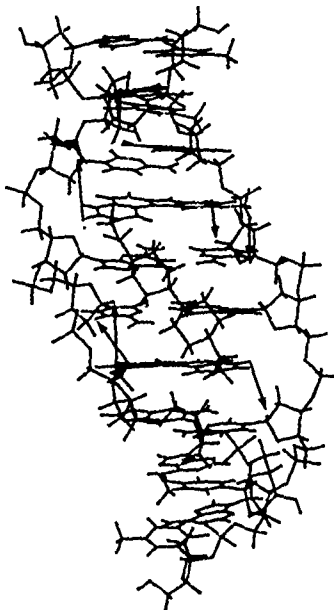
In the two octanucleotide complexes, inter-molecular NOEs were observed between drug aromatic protons H8 (and H9) and H1' (and H2', H2'') of cytosine 5, between drug H4 (and H5) and H1' of cytosine 3 (all these interacting in minor groove), and between drug O-CH₃ and H5 of cytosine 5 (major groove).

The inter-molecular observed NOEs clearly indicate major groove bis-intercalation (see figure 2).

Fig 2:

Preliminary model
of the ditercalinium
- $d(\text{TTCGCGAA})_2$
complex

↑ Intermolecular
NOEs



³¹P Resonances assignments

The phosphorus signals were assigned through 2D ³¹P {¹H} heteronuclear correlation experiments in the free nucleotides, whereas the bound phosphorus atoms were assigned by 2D - ³¹P exchange experiments.

(i) Octanucleotide complexes. In the octanucleotide complexes - the only ones with which 1:1 drug to helix ratio could be obtained - the signals in the analogue

complex were much broader than in the ditercalinium complex (fig 4). 2D - ^{31}P exchange experiments were carried out on all complexes (ratio 1:2 drug-nucleotide), but the analogue-octanucleotide complex did not show exchange between the ^{31}P in free and bound nucleotide even at 315°K, suggesting a longer lifetime in the nucleotide (in agreement with greater K_{affinity} measured with DNA). Thus, only the ditercalinium-octanucleotide complex ^{31}P resonances were unambiguously assigned. The two phosphorus atoms at the intercalation site CpG were downfield shifted (0.5 ppm and 0.9 ppm), as was the "inter-site" phosphorus GpC (0.5ppm). These results agree with previous results obtained on different nucleotides (ref 3). In the analogue complex, the two "site" phosphorus atoms are probably the two downfield shifted signals (as in ditercalinium complex), but the "inter-site" phosphorus is less shifted.

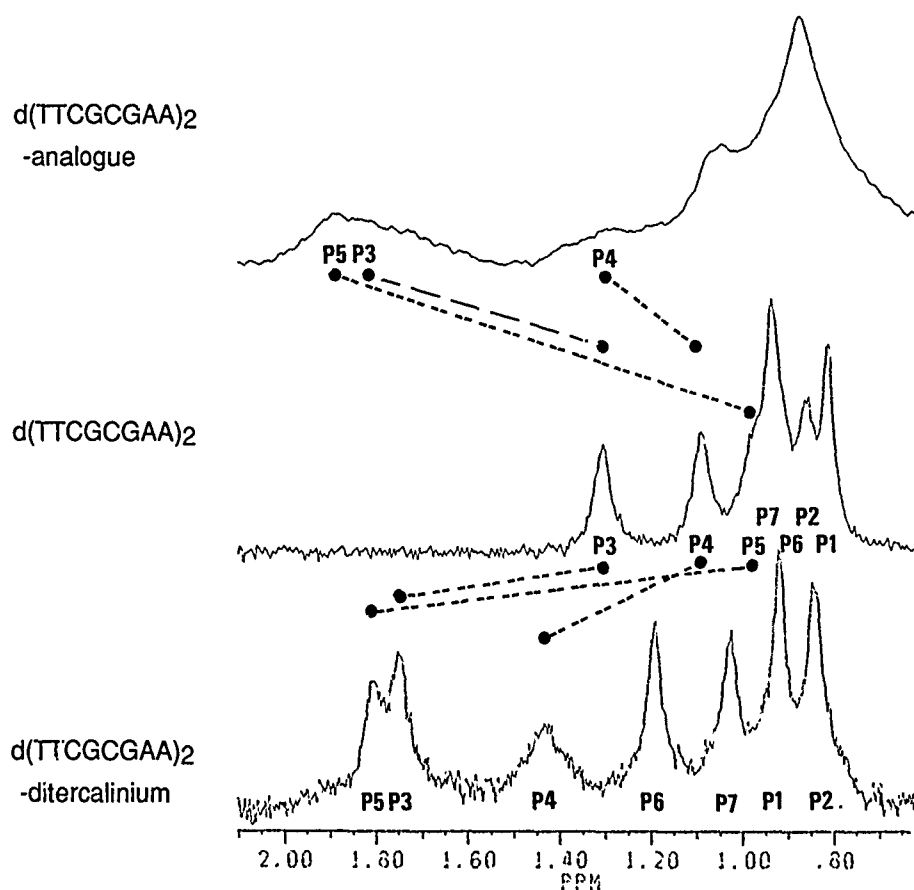


Figure 4: ^{31}P Spectra

(ii) Tetranucleotide complexes. Similarly, in the complexes formed with the two drugs and the tetranucleotide $d(CpGpCpG)_2$, the phosphorus signals at the intercalation sites were downfield shifted. Here the analogue complex showed exchange, allowing assignment of bound phosphorus atoms. The "inter-site" phosphorus in the analogue complex gave two bound signals, one upfield shifted and one downfield shifted, indicating two complexes or an asymmetrical complex at this level. This did not appear in the ditercalinium complex, where the "inter-site" phosphorus is only downfield shifted.

DISCUSSION

These results indicate that both drug dimers bis-intercalate into DNA via the major groove with similar geometry at the intercalation site, but with slight differences at the DNA level, particularly in the sugar-phosphate backbone. The question remains as to whether the difference in chemical shifts between ditercalinium and analogue complexes is due to a different geometry or/and different dynamics. To answer this question and to try to explain the difference of activity of these two drugs (at the molecular level), molecular modelling calculations will be carried out on these two complexes, and the relaxation matrix analysis of the models will be compared to the experimental 2D-NOESY spectra already obtained.

REFERENCES

- 1 D. Pelaprat, A. Delbarre, I. Le Guen, B.P. Roques and J.B. Le Pecq, DNA intercalating compounds as potential antitumour agents. *J. M. Chem.*, 23(1980) 1336-1343.
- 2 C. Garbay-Jaureguiberry, C. Esnault, M. Delepierre, P. Laugâa, S. Laalami, J.B. Le Pecq, J.B. and B.P. Roques, Rational Design of Bis-intercalating drugs as antitumour agents: importance of rigidity in the linking chain, *Drug Exptl. Clin. Res.*, XIII(6) (1987) 353-357.
- 3 A. Delbarre, M. Delepierre, C. Garbay-Jaureguiberry, J. Igolen, J.B. Le Pecq and B.P. Roques, Geometry of the antitumour drug ditercalinium bisintercalated into $d(CpGpCpG)_2$ by 1H NMR. *Proc. Natl. Acad. Sci. USA*, 84 (1987) 2155-2159.
- 4 A. Delbarre, M. Delepierre, B. Langlois d'estaintot, J. Igolen, and B.P. Roques, Bisintercalation of ditercalinium into a $d(CpGpCpG)_2$ minihelix: structure and dynamics aspects - a 400 MHz 1H NMR study, *Biopolymers*, 26 (1987) 1001-1033.

STRUCTURE DETERMINATION FROM NMR USING A RELAXATION MATRIX APPROACH; APPLICATION TO THE SOLUTION STRUCTURE OF CRAMBIN.

J.A.C. Rullmann¹, R.M.J.N. Lamerichs¹, C. Gonzalez², T.M.G. Koning¹, R. Boelens¹
and R. Kaptein¹

¹Department of Chemistry, University of Utrecht, Padualaan 8,
3584 CH Utrecht (the Netherlands)

²Instituto de Estructura de la Materia, C.S.I.C., Madrid (Spain)

SUMMARY

The structure of crambin in solution is determined from 2D NMR data. Distance constraints are obtained from nuclear Overhauser effect (NOE) measurements using an iterative relaxation matrix approach (IRMA), which is applied to a protein for the first time. *R*-factor calculations are performed on NOE buildups to describe the quality of agreement between theory and experiment. The final structure is within 1 Å (backbone r.m.s.) from the crystal structure. Refinement procedures are discussed.

INTRODUCTION

Crambin is a small, water insoluble protein of 46 residues, which can be obtained from the seeds of *Crambe abyssinica*. Its precise function is not known but the sequence is homologous to a number of plant toxins. The crystal structure was solved to 1.5 Å resolution (ref. 1); a 0.9 Å structure has also been obtained but the co-ordinates have not been published. Despite its small size the molecule contains two α -helices and a short β -sheet. Its three disulfide links are probably responsible for the high thermostability.

The present NMR study was undertaken for several reasons. First we wanted to establish whether there are appreciable differences between crystal and solution structures. Secondly, crambin with its variety of secondary structure elements and stable structure provided a good system to test a number of novel procedures; in developing these the crystal structure could be used as as reference point.

EXPERIMENTAL

Crambin displays amino acid heterogeneity at positions 22 (Pro or Ser) and 25 (Leu or Ile). It has been shown that the mixture as obtained in the isolation procedure consists of two species, the Ser/Ile and the Pro/Leu forms in a 55:45 ratio (ref. 2). The present analysis is based upon the NMR data for the Pro/Leu form. However, the crystal structure was determined assuming that the Pro/Ile form is the dominant component (ref. 1). We therefore replaced Ile-25 by Leu in the structure as obtained from the Brookhaven Protein Data Bank, and applied 30 steps of unconstrained energy minimization. The r.m.s. position change of the backbone atoms was only 0.06

Å. In the following this modified structure will be denoted as "the X-ray structure".

Two-dimensional NMR spectra of crambin in D₂O/acetone and in H₂O/acetone were recorded at 500 MHz. Proton resonances were assigned using the established sequential assignment procedures (ref. 3). The secondary structure has been reported before (ref. 4). It is identical to that in the crystal: α -helices are formed by residues 7-18 and 23-30, while the last residues of the first helix adopt a 3_{10} conformation; an antiparallel β -sheet is formed by residues 1-4 and 32-35.

Structure calculations were based upon 2D NOE spectra, in which 775 cross peaks could be assigned. The majority of these (646) had a buildup of sufficient quality to be used in a quantitative analysis. Spectra were recorded at six different mixing times ranging from 20 to 250 ms. All peaks were integrated by summing the intensity within a rectangular area around the peak.

ITERATIVE RELAXATION MATRIX APPROACH

Distance bounds were calculated from the observed NOE buildup curves by the Iterative Relaxation Matrix Approach, or IRMA (ref. 5). In this procedure spin diffusion is accounted for by solving the Bloch equation describing multispin relaxation

$$A(t_m) = A(0) \exp(-Rt_m) \quad (1).$$

where $A_{ij}(t_m)$ is the NOE intensity belonging to the spin pair (ij) at mixing time t_m , and R is the relaxation matrix. Starting from a model structure theoretical NOE values are calculated assuming that the molecule is rigid and rotates isotropically with a correlation time t_c . Theoretical values are replaced by experimental ones when available, and the combined NOE matrix is transformed back to a corrected relaxation matrix from which new distances are calculated assuming the same motional behaviour as before. Distance bounds are directly related to the variation of the back transformed relaxation matrix elements with the mixing time. Structure calculations are then performed using the new distance bounds. The whole process may be repeated starting from the optimized structure until convergence is obtained.

Relaxation matrix elements are calculated in the rigid rotor model directly from the interproton distances and the rotation correlation time (ref. 6). An extra decay term is added to the dipolar contribution on the diagonal to describe external leakage. We set this extra term to 1.25 s^{-1} . The correlation time t_c was set somewhat arbitrarily at 1 ns. Since R_{ij} is to first order linear in t_c , the scaling of experimental to theoretical NOE values makes the whole procedure rather insensitive to the exact choice of t_c as long as $\omega t_c > 1$.

Modifications are made to the relaxation matrix in order to describe the effect of methyl group rotations and of aromatic ring flips. One possibility is to add a kinetic constant to the cross-relaxation terms involving two mutually exchanging protons. This turns out to be equivalent (ref. 7) to a simple averaging of all corresponding matrix

elements of the equivalent protons, a procedure proposed by Landy and Rao for multiple-spin systems undergoing chemical exchange (ref. 8). The averaging method has been used here.

STRUCTURE CALCULATIONS

Computational procedures

In the first cycle calculations were performed with the Distance Geometry algorithm, derived from the original EMBED program (ref. 9). After embedding all structures were subjected to 200 steps of constraint function minimization. This was followed by two times 300 steps of Distance bound Driven Dynamics (DDD), i.e. a Newtonian dynamics simulation using chiral and distance constraint functions without force field terms (ref. 10). The first run was performed at 300 K; the second-run at 1 K, which has the effect of minimizing the constraint errors. All information about the covalent structure, i.e. standard bond lengths and angles, is represented by upper and lower bounds between atoms separated by less than four bonds.

Further optimizations were performed with Energy Minimization (EM) and Molecular Dynamics (MD) techniques using the GROMOS force field and programs (ref. 11). Since no solvent was included all net charges of side chains were reduced to zero. A cut-off of 8 Å was applied, and the pair list was updated every 10 steps. MD calculations were performed with the leapfrog algorithm, a 2 fs time step and with coupling to a heat bath using a time constant of 10 fs.

IRMA distance limits (ref. 5) were used directly as upper and lower distance constraints in the simulations. Pseudo atoms were introduced on all prochiral centres and the bounds were relaxed accordingly (refs. 3,12). For methyl groups pseudo atoms were used without correcting the bound: the process of kinetic averaging leads to a distance value that is more representative of the distance to the geometric centre of the three equivalent protons, i.e. the pseudo atom, than of the distance to any of the individual protons.

IRMA cycles

Two cycles of IRMA have been performed, starting from a fully extended chain. In the *first cycle* 8 DG-DDD structures were calculated, of which 7 belonged to one class; for these 7 the pairwise r.m.s. deviations for the backbone was 1.6 Å on average. One structure differed more than 6 Å from the others, but this one also had the largest violations. Before applying the Distance bound Driven Dynamics algorithm the average backbone r.m.s. value for the 7 structures amounted to 1.2 Å, which demonstrates that the DDD procedure improves the sampling of allowed configurational space. The best DG-DDD structure was simulated for 20 ps with restrained MD followed by EM. The result of this was used to recalculate the distance bounds with IRMA, starting the *second cycle*.

Using the new bounds structures were calculated in two different ways. The first procedure consisted of extensive MD refinement following an annealing strategy. The second procedure was a combination of DG-DDD with short MD optimization. The annealing simulation started from the best structure obtained in the previous cycle. First 25 ps of restrained MD were performed at 600 K. The elevated temperature increases the rate of transitions between allowed conformations. This is illustrated by the occurrence of a trans to cis peptide transition for Pro-19 after approximately 15 ps. Five structures from the trajectory, one after each 5 ps, were then simulated for 35 ps at 300 K. Five average structures were calculated from the last 20 ps of the latter trajectories and subjected to EM.

In the DG-DDD-MD procedure 10 structures were calculated with the DG-DDD technique (as described above) using the new bounds. To each of these 10 ps of restrained MD were applied after energy minimization. The final structures were again energy minimized.

RESULTS AND DISCUSSION

Annealing run

The potential energies and distance restraint energies of the five structures are remarkably similar, see Table 1, despite the conformational change of Pro-19.

TABLE 1

Total potential energies, distance restraint energies and total bound violations after MD and EM optimization (second IRMA cycle, annealing runs)^a.

Structure	Potential energy ^b kJ·mol ⁻¹	Distance restraint energy ^c kJ·mol ⁻¹	Total violation Å
A1	-2104	199	22.8
A2	-2079	200	23.2
A3	-2072	204	23.5
A4	-2120	207	24.0
A5	-2150	205	23.0

^a Data for minimized average structures from last 20 psec. of 5 different runs (see text).

^b Excluding distance restraint energy

^c Force constant 40 kJ·mol⁻¹·Å⁻²

The r.m.s. position differences averaged over the backbone atoms are 0.6 Å or less. The structures have converged with respect to the configurations obtained at 600 K: the backbone r.m.s. deviations between the latter five structures are around 1.0 Å. Overall the structure has converged to the X-ray structure: backbone r.m.s. differences with respect to the crystal structure have changed from 1.5 via 0.9 to 0.8 Å after completion of, respectively, the DG-DDD calculation, the first cycle and the second cycle. The largest differences between calculated and crystal structures occur in the

loop region between the helices and in the ten terminal residues.

When the r.m.s. differences are calculated over all atoms the five MD structures are within 1.3 Å (1.0 for the structures in which Pro-19 is trans), indicating that also most side chains are rather well constrained. All-atom r.m.s. positional differences with the X-ray structure are between 1.5 and 1.7 Å.

DG-DDD-MD

Three of the ten structures have distance restraint energies and total violations of similar magnitude as obtained in the annealing run. The other structures satisfy the constraints less well. However, the total potential energies (excluding the constraint energy) are all higher than for the annealed structures. The difference is at least 150 kJ·mol⁻¹, which far exceeds the thermal fluctuation.

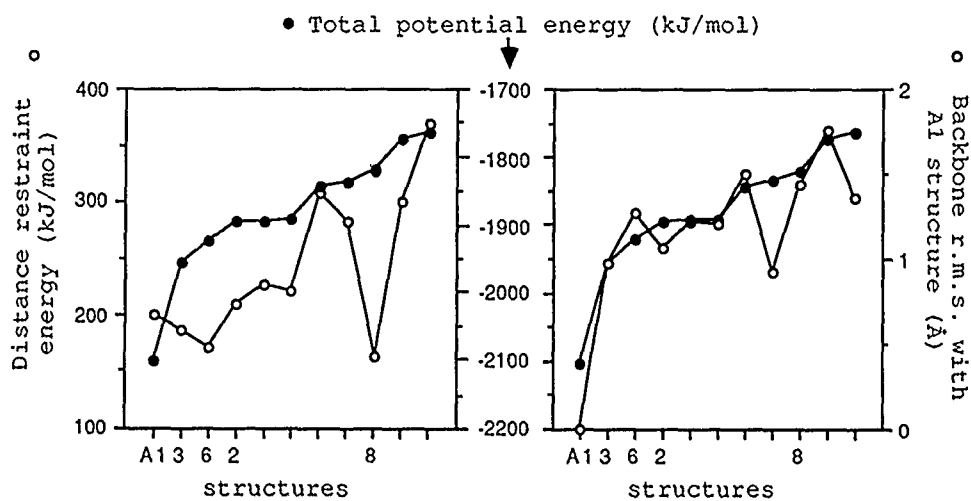


Fig. 1a (left) and 1b (right). Correlations between total potential energy (excluding distance restraint energy), distance restraint energy and r.m.s. positional difference. Data are shown for 1 structure obtained in annealing simulations (denoted as "A1") and for 10 structures obtained in DG-DDD-MD simulations (of which 4 are indicated by numbers), cf. notation in Tables 1 and 2.

The total energy is plotted against the distance restraint energy in Fig. 1a. Overall the quantities are correlated rather well (with the exception of one structure), showing that the force field and the NMR constraints are not inconsistent. The total energy is also correlated with the backbone r.m.s. positional difference with respect to one of the

annealed structures, see Fig. 1b. The relatively quick DG-DDD-MD procedure thus appears to probe around the minimum found by the more elaborate annealing procedure, yielding best structures that are approximately 1 Å (backbone r.m.s.) from the annealed structures.

R-FACTOR

May be the best way of judging the quality and accuracy of the present results is to compare directly measured NOEs with values predicted from the structure on the basis of eqn. (1). A useful definition is that of the *R*-factor

$$R = \frac{\sum_{i,j} \sum_{t_m} w_{ij} |A_{ij}^{\text{exp}}(t_m) - A_{ij}^{\text{th}}(t_m)|}{\sum_{i,j} \sum_{t_m} w_{ij} A_{ij}^{\text{exp}}(t_m)} \quad (2)$$

which is analogous to the crystallographic definition. Here the mixing times are used as weight factors, since the measurements are least accurate for the smallest mixing times. The summations in eqn. (2) are restricted to the interresidue NOEs since these are structurally the most important ones. In cases involving prochiral centres the theoretical NOEs are averaged over the two assignments. Table 2 shows that the *R*-factor has dropped considerably in going from the initial linear chain to the result of the second IRMA cycle.

TABLE 2

R-factors for different crambin structures^a.

Structure ^b	all	backbone - backbone	side chain - side chain
linear chain	0.99	0.92	0.94
after 1st cycle	0.56	0.39	0.66
A1	0.53	0.38	0.69
A2	0.56	0.36	0.91
A3	0.51	0.36	0.66
D2	0.52	0.36	0.75
D3	0.60	0.43	0.81
D6	0.56	0.35	0.67
D8	0.62	0.41	1.03
X-ray structure	0.47	0.41	0.56

^a *R*-factor, eqn. (2), calculated over interresidue contacts only.

^b Structures from annealing runs indicated by prefix "A", structures from DG-DDD-MD runs by prefix "D" (cf. numbering in Table 1 and Fig. 1).

The partial *R*-factor for backbone - backbone contacts, which is only indirectly influenced by the use of pseudo atoms, has dropped below the X-ray structure value, in agreement with the small but persistent structural deviation obtained in the simulations. The side chain - side chain partial *R*-factor exhibits much larger variation, reflecting the structural differences. However, the fact that one of the best annealed structures shows a particularly high, and the X-ray the lowest side chain - side chain *R*-factor, makes it clear that the present distance constraints are not sufficient to uniquely define the best structure. In particular the introduction of pseudo atoms on prochiral centres leads to a loss of precision. A second source of error is the neglect of local motions in the theoretical calculations.

Overall two of the annealed structures (A1,A3) and two of the DG-DDD-MD structures (D2,D6) appear to agree best with the experimental NOEs. Structure D8 which also had a low distance restraint energy but a relatively high GROMOS energy (cf. Fig. 1); shows a high *R*-factor due to improper positioning of the side chains.

CONCLUSIONS

The present constraint set for crambin in combination with the GROMOS force field corresponds to one well defined minimum, which as far as the protein backbone is concerned is close to the crystal structure. The residual violations are already quite small after the first cycle. Further optimization leads to a lowering of the GROMOS energy without any significant changes in the restraint energies. The best structures obtained from a combination of DG and short MD runs have constraint violations and *R*-factors that are similar to those in structures obtained in an extensive annealing optimization, but the total potential energy of the latter is significantly lower. The best structures differ by approximately 1 Å on the backbone from one another and from the X-ray structure. Side chain orientations are less well determined. We are currently investigating the effect of including in the simulations dihedral angle constraints obtained from J-couplings. The precision of the structure determination can further be increased by obtaining stereospecific assignments. A procedure to derive these assignments from a comparison between measured and calculated NOEs shows promising results.

REFERENCES

1. W.A. Hendrickson and M.M. Teeter, Structure of the hydrophobic protein crambin determined directly from anomalous scattering of sulfur, *Nature*, 290 (1981) 107-113.
2. J.A.W.H. Vermeulen, R.M.J.N. Lamerichs, L.J. Berliner, A. DeMarco, M. Llinàs, R. Boelens, J. Alleman and R. Kaptein, ¹H NMR characterization of two crambin species, *FEBS Lett.*, 219 (1987) 426-430.
3. K. Wüthrich, *NMR of Proteins and Nucleic Acids*, Wiley, New York, 1986.

4. R.M.J.N. Lamerichs, L.J. Berliner, R. Boelens, A. DeMarco, M. Llinàs and R. Kaptein, Secondary structure and hydrogen bonding of crambin in solution, *Eur. J. Biochem.*, 171 (1988) 307-312.
5. R. Boelens, T.M.G. Koning, G.A. van der Marel, J.H. van Boom and R. Kaptein, Iterative procedure for structure determination from proton-proton NOEs using a full relaxation matrix approach. Application to a DNA octamer, *J. Magn. Reson.*, 82 (1989) 290-308.
6. I. Solomon, Relaxation processes in a system of two spins, *Phys. Rev.*, 99 (1955) 559-565.
7. T.M.G. Koning, R. Boelens and R. Kaptein, in preparation.
8. S.B. Landy and B.D.N. Rao, Dynamical NOE in multiple-spin systems undergoing chemical exchange, *J. Magn. Reson.*, 81 (1989) 371-377.
9. T.F. Havel, I.D. Kuntz and G.M. Crippen, The theory and practice of distance geometry, *Bull. Math. Biol.*, 45 (1983) 665-720.
10. R. Kaptein, R. Boelens, R.M. Scheek and W.F. van Gunsteren, Protein structures from NMR, *Biochemistry*, 27 (1988) 5389-5395.
11. W.F. van Gunsteren and H.J.C. Berendsen, *Groningen Molecular Simulation (GROMOS) Library Manual*, Biomos B.V., Nijenborgh 16, Groningen, 1987.
12. J. de Vlieg, R. Boelens, R.M. Scheek, R. Kaptein and W.F. van Gunsteren, Restrained Molecular Dynamics procedure for protein tertiary structure determination from NMR data: a *lac* repressor headpiece structure based on information on J-coupling and from presence and absence of NOEs, *Israel J. Chem.*, 27 (1986) 181-188.

COMBINED USE OF NMR AND RESTRAINED COMPUTER SIMULATIONS FOR THE
MODELING OF THE CYCLIC ANTIFUNGAL LIPOPEPTIDE : STENDOMYCINE.

M. GENEST, B. STAWARZ, J.P. SIMORRE, D. GENEST, A. CAILLE AND M.
PTAK*.

Centre de Biophysique Moléculaire, CNRS et *Université d'Orléans,
1A Av. de la Recherche Scientifique, 45071 Orléans Cedex 2,
FRANCE.

SUMMARY

For determining the 3D structure of a molecule in solution, the most powerful method is based on measurements of proton-proton NOE's. But, their conversion in term of distances is ambiguous due to internal molecular motions and spin diffusion. We present an approach for the structure refinement based on the superposition of the experimental and calculated NOE's intensity which seems a better strategy than the fitting of the estimated and calculated distances. We show how to take account of internal motions of the molecule. This refinement method is applied to the conformational analysis of a cyclic lipopeptide : stendomycine.

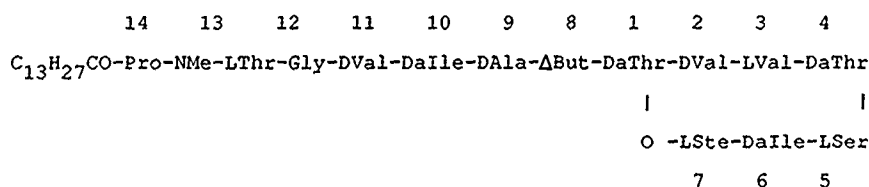
INTRODUCTION

3D structure determination of biological molecules from combined 2D NMR and computer simulations methods are essentially based on interproton distances derived from NOESY cross peak intensities measurements (refs. 1-2).

The NOE (Nuclear Overhauser Effect) intensities depend on the rates of cross relaxation of protons which in turn depend on fluctuations in the orientation and length of inter proton vectors ($1/r^6$ law) (ref. 3). In order to estimate distances which can be calibrated using a known distance (ref. 4) all inter proton vectors are assumed to move isotropically with identical correlation time. But a direct relation between cross peak intensity and cross relaxation rate neglects the uncertainties of internal motional behaviour of the molecule together with spin diffusion effect (ref. 5) induced by indirect magnetization transfer via other protons which leads to derive approximate ranges of proton-proton distances.

Questions about the reliability of the distances thus estimated led us to undertake NMR studies in order to analyse the internal dynamics of small molecules. Studies of NOESY cross peak intensities as a function of the temperature allow to discriminate between unvariable and fluctuating interproton distances and finally give a qualitative picture of the flexibility of the molecule. Furthermore, correlation times of fixed distances can be determined (ref. 6).

Such an approach has been applied to study the internal dynamics of the stendomycine molecule, an antifungal tetradecapeptide composed of a lactone ring and a terminal peptidic and fatty acid linear chain (ref. 7).



The reliability of the distances was expressed by different sets of weighting factors for distance constraints during the computer simulation process using the GROMOS program (ref. 8). A structural model of stendomycine is proposed and its validity is discussed on the basis of the comparison between the experimental and the theoretical NOESY maps (ref. 9).

METHODS

Molecular mechanics calculations were performed using the GROMOS potential energy functions, including bond stretching, bond angle bending, harmonic (out of plane, out of tetrahedral configuration) dihedral bending, sinusoidal dihedral torsions, Van der Waals and electrostatic interactions. A constraint potential term was added in order to let the molecule satisfy the set of NOE distance constraints : $V_{dc} = 1/2 K_{dc} (r_{ij} - r_{ij}^*)^2$, where r_{ij} is the actual distance between atoms i and j and r_{ij}^* the given distance constraint for this pair.

Chiral centers were specified along with planarity constraints for carbonyl carbons and amide nitrogens. All peptide bonds were set trans and the ester bond was kept near planarity.

No information was obtained about the C β chiral center of the

special residue stendomycidine, then the two possible configurations R and S have been examined. We focuss essentially our study on the structure of the lactone ring and on the folding mode of the beginning of the peptidic linear chain. The lack of information about the end of this chain prevents us to model its conformation correctly. In consequence, the computer simulations were undertaken on the truncated molecule including the DAla9 residue. A terminal acetyl group simulates the D Ile10 residue.

RESULTS and DISCUSSION

Starting structure

In this short paper we report preliminary results of the restrained energy minimizations of the model built structure derived from graphic manipulations on an Evans and Sutherland PS390 display system . The major difficulty is to build a graphic model which respects most of the NMR estimated distances together with an approximate ring closure.

Inter proton distances deduced from NOESY cross peak intensities can be divided in three groups (ref. 7). In the first one, distances are classified as reliable. In the second group the distances estimated from the $1/r^6$ approximation are certainly erroneous due to internal flexibility of the molecule. And in the last group, the distances are obtained from weak values of cross peak intensities. In this case, some of the distances are certainly correct and others are not.

Among the eight non fluctuating distances , the $H\beta$ (But8)- $H\gamma$ (But8) evaluated at 0.268 nm was used as reference for calibration. Such a set of distances involves mainly intra residue information except two of them (NH(But8)-NH(Thr1) and NH(But8)- $H\alpha$ (Ala9)) which lead to a particularly well defined structure of the beginning of the peptidic linear chain. Information such as NH(But8)-NH(Val2) proximity and detectable cross peak between NH(But8)- $H\alpha$ (Val3) tend to prove the proximity of one part of the linear chain with the lactone ring. Model built structure reflects this proximity, but the lack of reliable interproton distance is a strong handicap to ensure the unicity of the starting structure. However, other interproton proximities characterized by fluctuating distances unable us to

guide the structural arrangement of the residues.

Refined structure

First, strain in the graphic model was relaxed by performing steepest descent minimizations, and then series of constrained energy minimizations using the conjugate gradient algorithm followed by complete relaxation were applied successively.

Two approaches have been tested to run restrained energy optimizations. The first one does not take account of the reliability of the inter proton distances and all the constraints were equally weighted (force constant = 10^4 KJ mol⁻¹ nm⁻²) (Structures AR and AS). The second one considers weighting factors dependant on the reliability of the distances and a ratio of 10^2 was taken between force constant for reliable and unreliable distances (Structures BR and BS).

Comparison between constrained and total energies of the respective relaxed structures shows the BR one as preferential ($E_{dc}=144.5$ KJ mol⁻¹, $E_{Tot}=-108.5$ KJ mol⁻¹). Stendomycine including a S configuration for the Ste7 residue is relatively more strained (BS: $E_{dc}=152.4$ KJ mol⁻¹, $E_{Tot}=-54.4$ KJ mol⁻¹).

Superposition of the relaxed and constrained BR structures shows a good similarity. The relaxation of reliable distances concerns principally the distances involving the NH(Thr1) group. This corresponds to a favorable rearrangement of polar groups improving hydrogen bond interactions between the first residues of the peptidic linear chain and the Val residues of the lactone ring. Results obtained from both approaches in the management of the constraints are illustrated in figure 1 showing the superposition of the BR and AR relaxed structures. The folding of the lactone ring looks globally the same but the hydrogen bond network is completely different. For example, the CO(Ala2) is hydrogen bonded to the NH of the two Val residues in the BR structure whereas it is replaced by the CO(But8) in the AR form. In this last structure, the NH(Val2) tends to share this hydrogen bond with the CO(Ser5) carbonyl group.

Generally, structures optimized with equally weighted constraints led to strained geometries (high bond and angle energies), and chirality distortions, in order to satisfy at best the set of constraints. This result is illustrated in

Table 1 in which violations of reliable distances and strained energies are given for the constrained AR, AS, BR and BS structures.

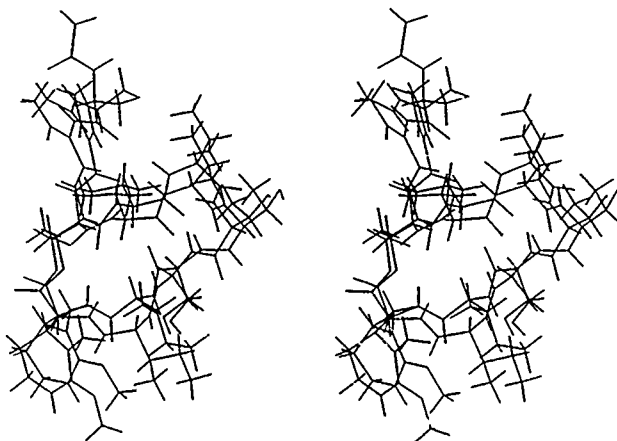


Fig. 1. Stereoview of the superposition of the relaxed structures, model of the stendomycine.

TABLE 1

Distances violations of the reliable interproton distances along with strained energies of the constrained structures AR, AS, BR and BS (see text).

Reliable distances	Distance violations (nm) for structures			
	AR	AS	BR	BS
NH (Ala9)-CaH (ALA9)	0.034	0.034	0.030	0.029
C H (Ala9)-NH (But8)	0.013	0.028	0.020	0.004
NH (But8)-NH (Thr1)	-0.017	0.023	0.012	0.011
NH (Thr1)-C β H (Thr1)	-0.058	0.031	-0.016	0.047
CaH (Val2)-C β H (Val2)	0.049	0.048	0.047	0.047
NH (Val3)-CaH (Val3)	0.001	0.001	0.013	0.013
NH (Ser5)-CaH (Ser5)	0.056	0.052	0.017	0.016
Strained energies (KJ mol ⁻¹)				
Bond	65.2	77.3	24.6	27.8
Angle	228.8	249.7	54.4	64.1

Another helpful criteria for discussing the validity of the refined structures is to compare the experimental and the

theoretical NOESY maps. Figures 2a and 2b show respectively the reconstructed NOESY experimental map exhibiting only the cross peaks with measurable intensities and the complete NOESY theoretical map relative to the BR relaxed structure. For clarity, all diagonal peaks are suppressed and in the experimental map many cross peaks are removed because they could not be attributed owing to strong overlap.

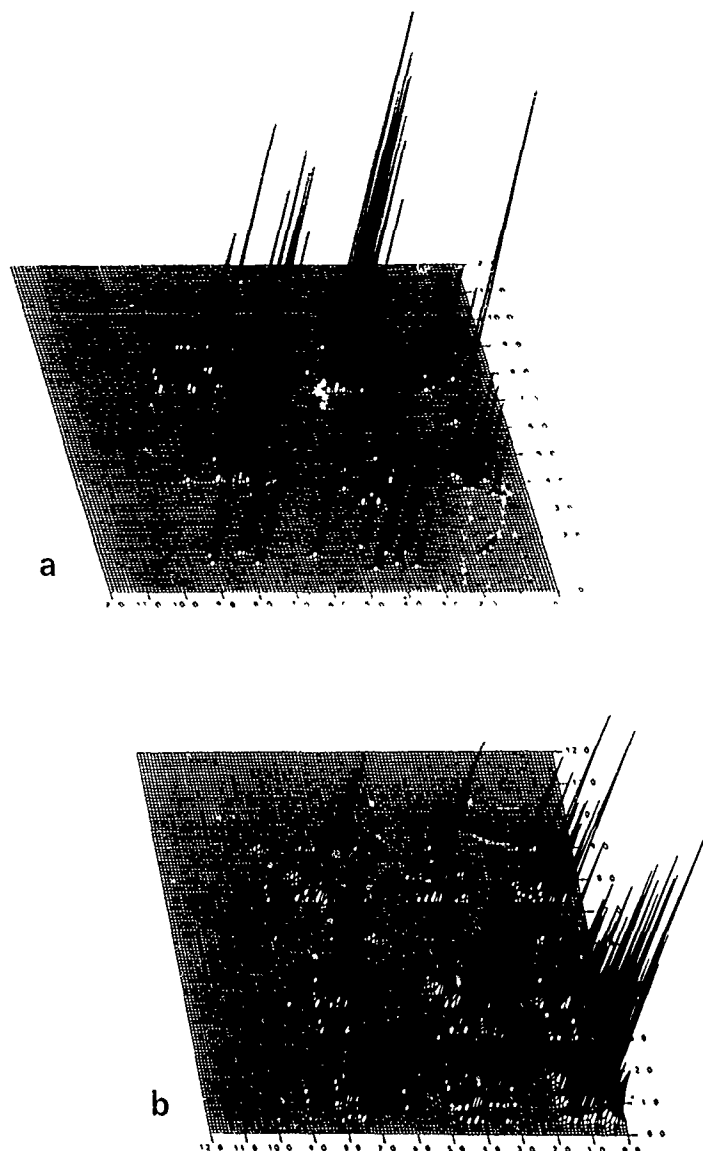


Fig. 2. Representations of the reconstructed NOESY maps. (diagonal peaks are suppressed) : a) experimental map. b) theoretical map relative to the best calculated structure modeling stendomycine.

A detailed comparison of all cross peaks seems difficult in view of the complexity of the theoretical map, but regions corresponding to the NH resonances call for comments. At the left top of the 2D map, The NH-NH cross peaks pattern shows noticeable differences in relative intensities values, while the NH-H α and the NH-H β , H γ cross peak distributions seem comparable. These observations are the consequence of internal motions. From a simple comparison between calculated and experimentally estimated distances it is not possible to evidence the influence of internal motions and we show here the utility of fitting two dimensional NOE spectra as already mentioned (refs. 9-10).

CONCLUSION

Two essential points emerge from this preliminary study of the structure modeling of the antifungal lipopeptide stendomycine.

First, NMR distance information have been used taking account on the flexibility of the molecule . It is well known that the NOE intensities depend both on interproton distances and on internal dynamics. From NMR studies, we were able to discriminate between constant distances which can be used quantitatively in molecular modeling and fluctuating distances which have to be used only qualitatively. For these last distances, the calculated NOE intensities using the classical formula are obviously different from the experimental ones owing to the $\langle 1/r^6 \rangle$ averaging.

Second , the role of the weighting factors introduced to manage distance constraints during the optimization process was to describe the reliability of the interproton distances estimated from 2D NMR. This strategy is somewhat different from the one usually applied where weighting factors increase with the NOE intensities . Such an approach is doubtful since it has been shown that weak NOE intensities are not necessarily related to large distances, and inversely strong NOE's can reflect distances larger than those estimated in the usual way (ref. 11).

REFERENCES

- 1 K. Wüthrich, NMR of proteins and nucleic acids, John Wiley

- and Sons, New York, 1986.
- 2 R.R. Ernst, G. Bodenhausen and A. Wokaun, Principles of nuclear magnetic resonance in one and two dimensions, Clarendon Press, Oxford, 1987.
 - 3 I. Solomon, Relaxation processes in a system of two spins, Phys. Rev., 99 (1955) 559-565.
 - 4 J.H. Noggle and R.E. Schirmer, The Nuclear Overhauser effect, Academic Press, New York, 1971.
 - 5 A. Kalk and H.J.C. Berendsen, Proton magnetic relaxation and spin diffusion, J. Magn. Reson., 24 (1976) 343-366.
 - 6 D. Genest and J.P. Simorre, A method for evaluating the reliability of distances and rotational correlation times deduced from 2D NMR NOESY experiment, Magn. Reson. Chem., in press.
 - 7 J.P. Simorre, D. Genest, A. Caille and M. Ptak, 2D NMR study of the internal molecular dynamics of the antifungal peptide stendomycine, submitted for publication.
 - 8 W.F. van Gunsteren, GROMOS program system, BIOMOS biomolecular software, Lab. of Physical Chemistry., Univ. of Groningen, 1986.
 - 9 D. Marion, M. Genest and M. Ptak, Reconstruction of NOESY maps. A requirement for reliable conformational analysis of biomolecules using 2D NMR, Biophys. Chem., 28 (1987) 235-244.
 - 10 E.T. Olejniczak, R.T. Gampe, Jr, and S.W. Fesik, Utility of fitting two dimensional NOE spectra, J. Magn. Reson., 81 (1989) 178-185.
 - 11 D. Genest, A Monte Carlo simulation of the influence of the internal motions on the molecular conformation deduced from 2D NMR experiments, Biopolymers, 28 (1989) 1903-1911.

MOLECULAR DYNAMICS STUDIES OF PRISTINAMYCIN II_A : FROM CRYSTAL STRUCTURE TO CONFORMATION IN APOLAR SOLVENT USING NMR DATA

E. SURCOUF¹, I. MORIZE¹, D. FRECHET¹, M. VUILHORGNE¹, A. MIKOU², E. GUITET²
& J.Y. LALLEMAND²

¹Rhône-Poulenc Santé, Centre de Recherches de Vitry, 13 Quai Jules Guesde, BP14, 94403
Vitry-sur-Seine (France)

²CNRS-ICSN, 91190 Gif-sur-Yvette (France)

SUMMARY

Pristinamycin (RP7293), a natural antibiotic, comprises two groups (I and II) of synergic compounds. Pristinamycin II_A is the major component of PII. The conformation of this molecule in apolar deuteriochloroform solution has been studied using quantitative nuclear Overhauser effects (NOEs) obtained by 1D and 2D NMR spectroscopy at 400 MHz and 250 MHz. Restrained molecular dynamics calculations have been performed using the INSIGHT/DISCOVER software package. 25 proton/proton distance constraints were taken into account in the simulations. The results were analysed using various programs developed in our group and interfaced with the program MANOSK. The major conformation obtained by restrained molecular dynamics calculation is found to be very close to that of the crystal structure and it has the same intramolecular hydrogen bond.

INTRODUCTION

Pristinamycin (RP7293) is a natural antibiotic, isolated from cultures of a soil organism *Streptomyces pristinaespiralis*. Pristinamycin is an association of two groups of synergic components : 30 to 40% pristinamycin I (PI) and 60 to 70% pristinamycin II (PII).

The chemistry of pristinamycins has been recently reviewed (ref. 1). The mode of action and the pharmacology of these molecules have been extensively studied. Pristinamycins are mainly used in the treatment of staphylococcal infections.

PI and PII are chemically different. Pristinamycin I is a mixture of three peptidic macrolactones of six aminoacid residues. The crystal structure of the major component (80-95%) PI_A (RP12535) has been recently determined in our group (ref. 2) and we are performing NMR experiments (ref. 3) in order to model the molecular structure in solution.

Natural pristinamycin II is a mixture of two polyunsaturated macrolactones. The major component (90-97%) is pristinamycin II_A (RP12536). This molecule was found to be identical to virginiamycin M1, the X-ray structure of which is already known (ref. 4).

In this paper, we will present a structure of Pristinamycin II_A in apolar solvent. Starting from the X-ray structure, the model was subsequently derived from NMR data obtained in deuteriochloroform solution, using restrained molecular dynamics simulations and classical minimisation techniques.

A set of 25 proton/proton distance constraints was taken into account for the calculations

allowing to determine the 23 dihedral angle degrees of freedom.

METHODS

NMR experiments

All experiments were performed on a 2×10^{-3} M solution of Pristinamycin II_A in deuterated chloroform. Overhauser effects were measured by 1D and 2D NMR on Bruker Spectrometers AM400 and WM250. Mixing times varied between 50 and 250 msec in the case of 2D experiments and 100 and 1000 msec in the case of 1D experiments.

Internuclear distances were calculated by the ISPA (Isolated Spin Approximation) method (ref. 5). The distance between protons H11 and H13 was used as a reference. Because of the conjugation between the two double bonds, this distance is fixed at 2.23 Å.

The list of the 25 proton/proton distances, determined by the ISPA method and used for the simulations, is given in Table 1.

TABLE 1

Proton/Proton distance constraints r_{HH} and estimated errors Δr .

n	H - H	r_{HH}	Δr -	Δr +
1	11 - 20	3.3	-0.2	+0.5
2	11 - 8	2.5	-0.2	+0.2
3	11 - 9	2.6	-0.2	+0.3
4	11 - 9'	3.3	-0.2	+0.5
5	11 -(33)	3.7	-0.2	+0.7
6	20 - 10	3.5	-0.2	+0.6
7	20 - 13	4.3	-0.2	+0.8
8	20 - 14	5.2	-0.2	+1.2
9	20 -(24)	3.8	-0.2	+0.7
10	20 - 9	4.5	-0.2	+0.7
11	20 - 9'	3.4	-0.2	+0.5
12	20 -(33)	4.7	-0.2	+0.9
13	17 - 20	4.2	-0.2	+0.8
14	26 -(32)	4.5	-0.2	+0.9
15	26 -(30,31)	4.1	-0.2	+0.8
16	6 - 8	3.0	-0.2	+0.4
17	6 - 3	3.1	-0.2	+0.4
18	6 - 4	2.6	-0.2	+0.3
19	6 - 32	3.8	-0.2	+0.8
20	6 -(30,31)	4.2	-0.2	+0.8
21	5 - 8	2.4	-0.2	+0.2
22	5 - 3	3.3	-0.2	+0.5
23	5 - 9	3.3	-0.2	+0.5
24	5 - 9'	4.1	-0.2	+0.8
25	5 -(32)	3.4	-0.2	+0.5

n : NOE constraint number.
The distances are given in Å.

Computational procedure

All the calculations have been performed using the INSIGHT/DISCOVER package (ref. 6). The analysis of the results was performed using various programs, developed in our group (ref. 7) and derived from the MANOSK software (ref. 8).

The potential energy function has the classical analytical form, but this expression contains "off diagonal" or "cross terms". Those terms were not involved in our calculations, since too many values about non-peptidic internal coordinates are missing. For the same reason we used a simple harmonic potential for the bond stretching term of the energy, instead of a Morse potential.

In order to take into account the NOE distance constraints, it is possible with DISCOVER to use a special term which is a skewed biharmonic function of the form :

$$E_{res} = \begin{cases} C_1 (r-r_0)^2 & \text{if } r > r_0 \\ C_2 (r-r_0)^2 & \text{if } r < r_0 \end{cases} \quad (1)$$

with :

$$C_1 = kTS / 2(\Delta r^+)^2$$

and :

C_i : force constant

r_0 : target distance

T : absolute temperature

Δr^+ : positive estimated error

$$C_2 = kTS / 2(\Delta r^-)^2$$

k : Boltzmann constant

r : calculated distance

S : scale factor

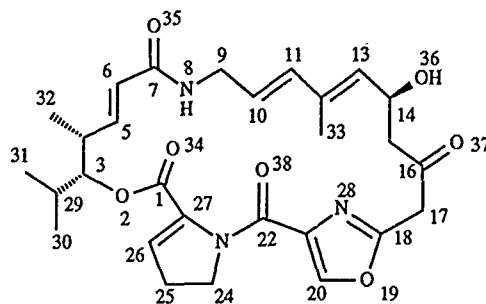
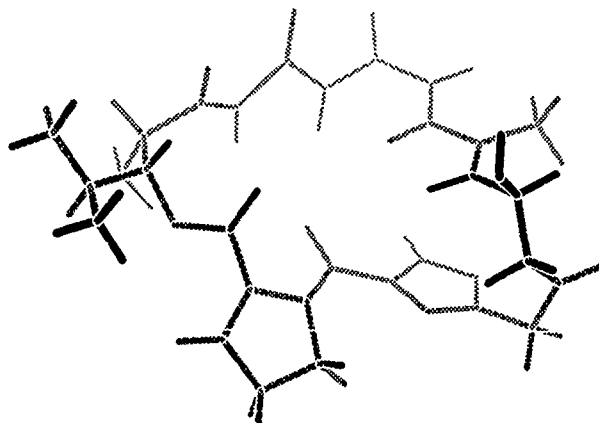
Δr^- : negative estimated error

All the 35 hydrogen atoms were treated explicitly in the calculations. However, when the NOE distance constraints refer to methyl or isopropyl group, using DISCOVER's special facility, we can define average atoms, the coordinates of which are the arithmetic average of all the hydrogen atoms of each group. In the case of PII_A, 4 average atoms were defined for CH₃ (32), CH₃(33), C(CH₃)₂ (30,31) and CH₂ (24).

The "Consistent Valence Force Field" (CVFF) from BIOSYM package was chosen (ref. 6) for dynamics calculations using DISCOVER. Some parameters were modified and the missing ones were determined and adjusted by using X-ray crystal structure minimisation techniques and by comparing with other force fields. Indeed, pristinamycin II_A, as shown in Fig. 1, contains an oxazole nucleus incorporated into the macrolactone structure, which is relatively uncommon. Therefore the determination of the force field parameters concerning this part of the molecule required special studies.

Starting from the conformation of PII_A in the crystal structure (Fig. 2), the structure was relaxed by performing an energy minimisation of 100 steps of steepest descent (SD) and 453 iterations of quasi Newton Raphson (VA09 minimizer in Discover) until a final energy of 76.8 Kcal mol⁻¹. This conformation is noted XRMIN.

Then, starting from XRMIN, the conformational space was searched using restrained dynamics calculations at high temperature, in order to generate various starting conformations for further restrained molecular dynamics simulations.

Fig. 1. Pristinamycin II_A.Fig. 2. Crystal structure of PII_A.

The following simulation procedure (RMD) has been performed 100 times :

- 2 ps (2000 steps of 1 fs) of dynamics at 900 K
velocity randomisation, temperature relaxation time of $\tau_t = 0.01$ ps
- 2 ps of dynamics at 300 K, $\tau_t = 0.1$ ps
- minimisation, storage of the conformation

All the NOE distance constraints were included in all the steps of the simulation with a scale factor (S) of 2, leading to a force constant of 15 Kcal mol⁻¹ at 300 K for a Δr equal to 0.2 Å (see equation 1). In this procedure, the conformation at the end of a minimisation is the starting conformation of the next step.

A set of 100 conformations was obtained (noted : RMDi, i=1,100) and several families (noted : Fj, j=1,n families) of conformations were derived after analysis.

Finally, for each family, starting from the most representative conformation, the following procedure (FRMD) was performed :

- 10 ps of dynamics at 300 K, velocity randomisation, $\tau_t = 0.01$ ps
- a restrained and continue dynamic run of 50 ps, $\tau_t = 0.1$ ps

Along the trajectory, every 0.5 ps a conformation was archived. So, a set of 100 snapshot structures per family (FjRMDi) was obtained for analysis. These snapshot structures were subsequently minimised with and without NOE distance constraints for a more realistic energy comparison with the previous calculations (RMDi) and with the crystal structure XRMIN.

During these last restrained dynamics calculations, to allow more flexibility to the molecule, the values $\Delta r+$ and $\Delta r-$ were at least of $\pm 0.5 \text{ \AA}$ and the NOE distance constraints corresponding to average atoms were not included.

All the calculations were performed on a FPS264 array processor connected to a VAX 8550 computer. The simulation of 1 ps of a PII_A molecule (73 atoms) took about 1 mn 12 s CPU time on the FPS.

RESULTS AND DISCUSSION

High temperature restrained dynamics : RMD

Figure 3 shows the evolution of the potential and total (potential + restraint) energies during the dynamic run. It is observed that the potential energy varies between 81.2 and 98.7 Kcal mol⁻¹ and the restraint energy fluctuates around 11.7 Kcal mol⁻¹. Several conformational changes occur during the simulation leading to different conformations. This illustrates one of the advantages of high temperature dynamics simulations for conformational space searching. Indeed, the kinetic energy allows the molecule to cross over barriers.

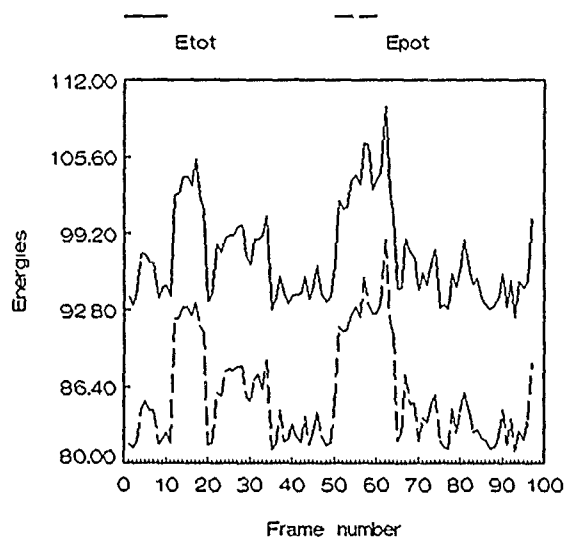


Fig. 3. Variation of potential and total energies during RMD procedure. The energies are given in Kcal mol⁻¹.

Statistics about NOE distance constraints for RMDi conformations are given in Table 2. It can be seen that, only 8 distance constraints are observed outside the ranges defined for the target distances. The largest average difference is 0.42 \AA . This indicates that all generated structures

are in good agreement with the set of constraints.

TABLE 2
Statistics on distance constraints for RMD procedure.

n	N _{out}	<dif>	d _{max}	d<r>
3	90	0.16	0.27	0.14
9	91	0.21	0.38	0.18
14	51	0.22	0.91	0.10
15	62	0.15	0.33	-
16	85	0.11	0.15	0.09
20	78	0.16	0.56	0.02
23	100	0.42	0.61	0.42
25	100	0.08	0.12	0.08

n : NOE constraint number.

N_{out} : number of conformations for which r is out of the range $R = [r_0 - \Delta r, r_0 + \Delta r]$.

<dif>: average of the differences between r and the nearest limit of R.

d_{max} : largest difference between r and the nearest limit of R.

d<r> : difference between the average of r and the nearest limit of R.

Distance values are given in Å.

The set of 100 RMDi conformations can be analysed in terms of rms values. A rms matrix was produced by superimposing each conformation RMDi on each other. All the superimpositions were performed by taking into account the heavy atoms only. In order to avoid the influence of the isopropyl group orientation on the rms values, the two terminal carbon atoms of this group were not included. The largest rms difference is 2.13 Å.

In order to determine correlations between conformations, a rms limit criterion can be used to classify them into families. Taking a rms limit of 0.4 Å, which is a reasonable value, we could deduce 4 families. Only 3 conformations remained not correlated to the others. The conformational change which occurs from one family to another is consistent with the variation of energy given in Fig. 3. The major family (F1) comprises 58 conformations. These conformations happen to have the lowest potential energy.

Finally, one conformation per family was selected to perform the subsequent analysis and calculations. Their characteristics are summarised in Table 3. It can be seen that conformations F1 and F4 differ only by a rms of 0.61 Å. However F3 deviates from F1 and F4 with a rms of about 1.1 Å and F2 is even more different.

Restrained dynamics at 300 K : FRMD

Since F4 is close to F1, the FRMD procedure was applied only on F1, F2 and F3 structures. Therefore 3 sets of 100 FjRMDi conformations were analysed using energy plots and rms matrix. In addition, "cross-rms matrices" have been produced by superimposing each conformation of Fj family on each conformation of Fk family.

It can be observed that, starting from three initial conformations which have a rms difference of 1.1 to 1.8 Å, after about 10 ps, all the FRMD calculations converge to the same final structure in the limit of 0.5 Å for the rms criterion. It should be noted that the structures are not minimised

during these procedures.

At last, a restrained minimisation procedure was applied to the three conformations F1RMD100, F2RMD100 and F3RMD100. The rms differences between the structures are very small (0.2 and 0.3 Å). A "cross-rms matrix" calculated by superimposing the initial conformations F1, F2 and F3 (before FRMD procedure) and the final conformations F1RMD100, F2RMD100 and F3RMD100 (after the restrained minimisation) shows that the three families converge to a conformation which is close to the major family F1 (table 4). Therefore we can conclude that the PII_A molecule adopts the conformation F1 in apolar solvent.

TABLE 3

Characteristics of the 4 families.

Fj	RMDi	E _{pot}	E _{res}	----	rms	----
				F1	F2	F3
1	35	81.2	11.6	-	1.45	1.14
2	12	92.1	10.4	1.45	-	1.81
3	64	90.7	9.7	1.14	1.81	-
4	50	85.2	11.3	0.61	1.62	1.15

j : family number.

RMDi : conformation number in the RMD procedure.

E_{pot} : potential energy.

E_{res} : restraint energy (25 NOE constraints).

The energies are given in Kcal mol⁻¹.

The rms differences are given in Å.

TABLE 4

Characteristics of the FjRMD100 conformations.

FjRMDi	E _{pot}	E _{res}	----	rms	----
			F1	F2	F3
1 100	79.4	6.9	0.69	1.19	1.57
2 100	81.3	6.9	0.76	1.22	1.63
3 100	81.6	6.7	0.63	1.26	1.52

FjRMDi : conformation number for each Fj family in FRMD procedure.

Fj : initial conformation of family number j.

E_{pot} : potential energy.

E_{res} : restraint energy (17 NOE constraints).

The energies are given in Kcal mol⁻¹.

The rms differences are given in Å.

Comparison of the solution and crystal structures

Comparing the conformation F1 with the X-ray structure XRMIN, we show that both conformations are very close. Indeed, the overall rms difference for the heavy atoms of the backbone (without the two carbon atoms of the isopropyl group) is about 0.55 Å. The largest deviations occur around the C5-C6 bond and around the hydroxyl group O36. This is consistent with the crystal structure features. As a matter-of-fact, in the crystal structure, the hydroxyl group is engaged in an intermolecular hydrogen bond and the measured length of the bond

C5-C6 is shorter (1.14 Å) than that of a normal covalent bond. This last observation was interpreted (ref. 4) as a rocking of C5-C6 in the plane defined by C4-C5-C6-C7.

In both structures XRMIN and F1, the orientation of the C5-C6 bond differs by about 180 degrees. So, in order to control the flexibility of PII_A in this part of the molecule, a dynamic run of 50 ps at 300K (without distance constraints) was performed, starting from the XRMIN structure. After a few ps, an inversion of the dihedral angle C4-C5-C6-C7 round the bond C5-C6 can be observed.

Finally, it should be noted that the intramolecular hydrogen bond between N8 and O38 (H...O = 2.1 Å) is observed both in solution and in the crystal. This hydrogen bond induces a compact and stable structure.

CONCLUSIONS

Starting from the crystal structure and from 25 proton/proton distance constraints determined by NMR experiments, and using restrained molecular dynamics simulations at high temperature, we derived three families of conformations which satisfy the set of NOE constraints. Then, dynamics simulations at 300 K, performed for one conformation in each family, converged towards a unique conformation. This conformation is very close to that observed by X-ray crystallography.

Pristinamycin II_A is a good test for the refinement of our strategy in molecular modelling calculations and for the choice of the force fields, since various experimental informations are available concerning this molecule: X-ray structure, NMR data, and also Laser Raman spectroscopy measurements in the solid state (ref. 9).

Therefore, we are pursuing the present work, with the investigation of different approaches allowing to get more accurate values of the force field parameters. The first one relies on "ab initio" calculation of the molecular orbital energy surface, using a method developed by the Biosym Force Field Consortium (ref. 10). The second one is based on the calculation of the vibrational frequencies and their comparison with the experimental Raman vibrational frequencies (ref. 9).

REFERENCES

- 1 J.M.Paris, J.C.Barrière, C.Smith & P.E.Bost, The Chemistry of Pristinamycins, to be published in "Recent progress in the Chemical Synthesis of Antibiotics", Eds G.Lukacs, M.Ohno, Springer Verlag, Heidelberg.
- 2 J.P.Leclerc, J.P.Mornon & E.Surcouf, in preparation.
- 3 E.Surcouf, F.Herman, D.Fréchet & M.Vuilhorgne, in preparation.
- 4 F.Durant, G.Evrard, J.P.Declercq & G.Germain, Cryst. Struct. Comm., 3, (1974), 503-510.
- 5 A.M. Gronenborn & G.M. Clore Prog. Nucl. Mag. Res. Spectr., (1985), 17, 1-32.
- 6 INSIGHT/DISCOVER, Biosym Technologies Inc., 10065 Barnes Canyon Road, San Diego, CA 92121.
- 7 I.Morize & E.Surcouf, in preparation.
- 8 MANOSK, J.Chérif, M.C.Vaney, I.Morize, E.Surcouf, N.Colloc'h & J.P.Mornon, J. Mol. Graphics, (1988), 6, 155-160.
- 9 G.Vergoten, personal communication.
- 10 J.R.Maple, U.Dinur & A.T.Hagler, Proc. Natl. Acad. Sci., USA, (1988), 85, 15, 5350-5354.

ARTIFICIAL INTELLIGENCE IN THE MOLECULAR WORLD

M.C. HATON

CRIN/INRIA, University of Nancy I, BP 239, 54506 Vandoeuvre-les-Nancy (France)

SUMMARY

Various aspects of AI may enter the molecular world : knowledge and reasoning representation, heuristic problem solving, planning, image and natural language understanding, knowledge-based and expert systems, "intelligent" computer-aided instruction, symbolic learning. These aspects will be discussed, focusing on the specific problems the chemist or physicochemist has to face with (synthesis, experimental planning, structure elicitation, spectra interpretation, classification, information retrieval, fault diagnosis in analytical instrumentation, student education and so on). Some examples, in a non exhaustive way, will be given.

INTRODUCTION

While solving problems is a long tradition in experimental sciences, research into how people actually solve that kind of problems is relatively recent. It was supposed twenty years ago that people knew how to perform problem solving and overall knew how to teach their solving methods.

Reflexion has been recently deepened in the frame of cognitive sciences, namely by knowledge psychologists and now by knowledge engineers in the frame of a branch of advanced information processing named "Artificial Intelligence" (AI). AI deals with these activities which characterize human behaviour : knowledge acquisition and structuration, reasoning, perception, decision making, etc. AI methods, tending to formalize human knowledge and reasoning processes, give us new tools for a better understanding of that processes.

Dealing with molecules is a complex process, which, paradoxically, is nowadays complicated by the fact that we can have more and more detailed representations and more and more numerous data concerning them. While the use of computers is increasing, the amount and the complexity of the data involved in the different domains of chemistry make it imperative to discriminate which information is relevant and to consider new ways of processing it. The chemist is often able to extract this relevant information and to make decisions thanks to his knowledge, his experience, his intuition and so on.

AI may help the specialists in these tasks, together with many classical methods of data processing.

WHAT IS ARTIFICIAL INTELLIGENCE ?

A definition

AI may be considered along two complementary axes : the first one deals with the study of the mechanisms of human intelligence, the computer being used as a tool of simulation to test a model or a theory, the second one, more pragmatic, is concerned with the efforts made to give to the computer some abilities usually associated to the human intelligence (knowledge acquisition and structuration, perception, reasoning, decision making, etc.). Most of the time, this second point of view is envisaged. It consists in emulating, with a computer program, some intelligent behaviours, without trying to reproduct the corresponding functioning of the human being. In that way, AI appears to be an advanced branch of computer science, but in connection with many other domains.

AI is presently a real science, both for academic reasons (several thousand researchers and teachers all around the world) and economic ones (well mastered applications, specialized companies, important research/development projects) {2}.

Main characteristics

Though the subjects of interest are various and numerous, there are some common features among the AI systems.

At first, they handle symbolic information together with numbers. This type of information represent concepts, rules, objects or facts that are similar to those used by the human being when reasoning. Non classical programming, like logic programming with PROLOG, functional programming with ML or LISP, object-oriented languages are devoted to this task. We may mention too the new trends in computer architectures issued from the study of human memory and reasoning.

Then, they use heuristic methods which can be opposed to the classical algorithmic ones. Heuristics allows reasoning in a non deterministic way, the success of which is not warranted, but which, when everything is all right, gives a good solution to the problem in question with time saving. In case of failure, it backtracks and tries another solution. The heuristic search of a solution often consists in "pruning" the set of the possible solving paths to only consider the best promising ones. The use of heuristic methods allows to deal with two types of problems that cannot be treated by classical methods :

- the problems for which we do not know any algorithmic solution. It is often the case inside the human activities like perception, decision making, design, ..
- the problems for which the solving complexity increases the available computing means. A typical example is the one of the "chess" game.

AI may also deal with missing or approximate data. It corresponds to usual situations, for instance in medicine, company management, and so on. Even if it does

not entirely solve these problems, AI provides efficient methods, like approximate or non-monotonous reasoning.

Another important feature of AI is the concept of knowledge. By opposition with the first researchers that tried to develop "general problem solvers" [3], AI presently designs systems able to solve problems in limited domains, taking into account a great amount of relative knowledge. Once more, this feature may be pointed out in the human activities : the human expertise relies on that type of knowledge together with experience, intuition, etc. The knowledge-based systems and the expert systems take a large place inside AI.

At last, we can say that AI needs to enter various domains, besides computer science : logic, cognitive psychology, linguistics, cybernetics, neurosciences, ...

The domains of AI

Briefly, we mention automatic theorem proving, natural language processing, automatic speech understanding, image interpretation and computer vision, robotics, games and expert systems.

KNOWLEDGE AND REASONING REPRESENTATION

Introduction

The word "knowledge" includes all the types of human knowledge : objects of the real world, facts and events, larger concepts corresponding to groupements or generalizations of basic ones, relations between concepts, heuristics and know-how strategies, reasoning procedures, etc. The word "metaknowledge" relates to the confidence that we associate to a piece of knowledge.

Knowledge may be problem specific, domain specific, shallow, deep, exact, approximate, .

Knowledge representation consists in translating it in a symbolic form that may be understood by a reasoning system.

A representation mode associates a data structure to represent the useful information and the procedure of manipulating it. The reasoning mechanism allows to dynamically discover new pieces of information or knowledge concerning the problem. This feature characterizes a knowledge base, compared to a data base from which it is only possible to extract pieces of information that have been explicitly introduced in it.

One opposes sometimes declarative and procedural knowledge. This duality corresponds to the distinction between a descriptive representation ("what") of the knowledge and the representation of know-how ("how to do"). The present trend is to develop mixed representation modes.

The knowledge representation techniques

They can be divided in several categories :

- the logical representations : they rely upon the proposition calculus (0-order logic) or the first order predicate calculus (1-order logic). The way of reasoning consists in deriving new formulas from the initial ones. For that purpose, a set of logical rules are available, like the "resolution" rule. Moreover, there exists non standard logics allowing to deal with ambiguous or fuzzy knowledge,

- the semantic networks, at first used for the interpretation of natural language. They are graphs in which the nodes represent concepts and the arcs represent semantic relations between these concepts. The way of reasoning is to match the network representing the problem with the knowledge network thanks to a pattern matching procedure. The matching is rarely immediate and often needs what is called "the inheritance of properties from a general concept to a more particular one,

- the production rules which are pieces of knowledge in the form :

if conditions then conclusions and/or actions (coefficient),

which means that, if the premises (condition side) are valid, it is possible to draw some conclusions, start an action, fire a computing or display procedure, etc. The coefficient, if it exists, indicates the confidence about the rule or its likelihood. It allows to manage uncertainty and comes generally from the experience of the expert.

The reasoning process is done by an interpreter in two possible ways :

- . the forward or data-driven chaining, considering the rules from the left to the right and drawing conclusions from the current data,

- . the backward or goal-driven chaining, which consists, while trying to prove the right part of a rule, to prove the elements of the left side.

These two chainings may be alternately performed in a mixed approach, well adapted to reproduce the human reasoning based at first upon data and then directed by hypotheses,

- procedural representations. They allow to introduce classical algorithmic procedures inside AI systems,

- the object-oriented representations. They come from the necessity to organize the set of available knowledge. Various formalisms have been developed (frames and scripts, prototypes, ...) that led to the concept of structured objects. These objects are now the basis of object-oriented programming languages.

HEURISTIC PROBLEM SOLVING - PLANNING

It is now accepted that the memorization and computing abilities of a computer are not sufficient to solve complex problems. In the case of complex games needing strategies, of planning (determination of the sequence of actions to be performed to reach a prerequisite goal), etc., it is necessary to use specific techniques depending

on different factors : the best solution is needed, backtracking is possible, we can forecast the effect of an action, the problem may be divided into sub-problems easier to solve or not, and so on. Different modes are available to represent the process of problem solving (state space trees, and-or trees). The virtual resolution trees are not entirely developed, they generally are enormous. Heuristics intervene to develop at best the branches that have some chance to lead to a good solution.

IMAGE INTERPRETATION

Computer vision aims at the interpretation of an image or a scene and the matching of this interpretation with an action upon the environment. A vision system includes modules performing image processing and pattern recognition. It is based on two main ideas : the need of new architecture for parallel computation, the existence of a hierarchy of representations, since the physical signals up to their cognitive interpretation.

NATURAL LANGUAGE (NL) UNDERSTANDING

Natural language, like french or english languages, are preference means of interaction between men and machines. Since the access to data inside a database only requires a formal artificial language, to be able to interact with the machine (to ask a question or to receive an answer or a reformed text) in a natural way constitutes an important step. NL understanding concerns both written and oral languages.

KNOWLEDGE-BASED AND EXPERT SYSTEMS

Introduction

The current success of expert systems (ES) mainly originates from their ability to carry out such tasks as diagnosis and decision making, which were little computerized.

The advent of ES in the early 70s corresponds to a global evolution of Artificial Intelligence from the search for general problem solving techniques to the study of the mechanisms used by a human expert to solve one problem in a given field. A pioneering project was the DENDRAL project, launched in 1965 by E. Feigenbaum and his colleagues at the University of Stanford. DENDRAL helps the chemist to infer the structure of an organic compound from its mass spectrogram and its rough chemical formula {4}.

This project clearly showed the interest of separating the knowledge necessary for solving a problem in a given field (e.g. organic chemistry) from the general mechanisms of knowledge handling. The idea of knowledge-based systems lies precisely in this separation. Several projects then completed this basic idea : MYCIN (Shortliffe, 1976, at the University of Stanford {5}) for the diagnosis of some viral blood

diseases, which greatly influenced the design of a large number of ES, MOLGEN (Stefik at the University of Stanford) {6} for the planning of experiments in molecular genetics, or PROSPECTOR (Duda and Hart, 1979, at the Stanford Research Institute) in the field of geology {7}.

Since that beginning, a large number of ES have been developed and routinely used. An expert system is made up of a set of computer programs that reach the performances of a human expert for a precise task in a restricted domain through the use of a knowledge base extracted from experts of the domain.

Second generation ES, based on a deep knowledge about the application field are now appearing.

Basic architecture of an expert system

An ES is a particular case of knowledge-based system. It basically includes three parts :

- one or several knowledge bases which constitute the long term memory of the system and which store, possibly with several formalisms, as seen previously, the various pieces of knowledge necessary for solving a problem : permanent facts, know-how, rules of thumb, common-sense knowledge, etc. The formalism of production rules has been widely used in ES since they are well suited for representing the way in which an expert solves a problem, especially in domains like diagnosis,

- a fact base which contains the data and facts related to a problem to be solved. It constitutes the short term working memory of the ES,

- an inference engine, or interpreter, in charge of reasoning about a problem by exploiting the facts and the knowledge available.

An inference engine works according to a particular search strategy. For simple cases, the strategy may only consist in exhaustively exploring all potential solutions. It is however usually necessary to design more sophisticated mechanisms, eventually based on meta-knowledge (such as the metarules used in MYCIN), in order to reach a solution in a reasonable time.

The explicit separation between the knowledge base and the inference engine is useful to implement practical knowledge-based applications (especially for the incremental design of large systems, the relative ease of updating and maintaining knowledge bases, etc.).

Most present systems use production rules for coding partly or entirely their knowledge bases. That offers several advantages :

- readability of knowledge,
- ease of building up the base since production rules are independent
- possibility for the system to explain its reasoning.

Object-oriented languages offer now a promising way of coding knowledge and incorporated reasoning mechanisms.

Typology of problems

ES can help solving problems with the following characteristics :

- a large body of knowledge, possibly incomplete and uncertain, is necessary for solving a problem,
- the application field may evolve in time,
- the resolution of a problem is basically heuristic,
- symbolic processing plays a major role in the problem solving process.

Problems tackled by ES can be divided into three main classes :

- data interpretation : the problem consists in examining a set of data and/or physical signals in order to interpret them. Typical examples are diagnosis (in medicine, industry, finance and banking, etc.) and signal and image interpretation (in biology, physicochemistry and medicine, acoustics, industry, etc.),
- planning : this domain concerns complex tasks that can be found in activities like decision making, administration and law, resource management, production management, robotics and so on,
- design : of objects (e.g. VLSI chips) according to preliminary specifications.

Finally ES are also used in computer-aided instruction (CAI) due to their ability of explicating and explaining their reasoning schemes. Applications exist in several domains : medicine (learning how to give a diagnosis), industry (teaching instructions to process controllers), academic teaching (for mathematics, physics, foreign language, etc.).

Domains of application

The different categories of ES that have just been presented can be found in very various domains of application. The following list is not exhaustive and includes systems at various stages of development (first demonstrators, prototypes, operational products) : agriculture, avionics and space, bank, finance and insurance, biotechnologies, chemistry, computer science, electronics, geology, industry , law and regulation, mathematics, medicine, military, teaching and so on.

Development of an expert system

The decision of launching the development of an ES must be taken carefully by taking into account several criteria :

- the problem to be solved necessitates to use not only quantitative but also qualitative information,

- there exist known experts of the domain who are motivated and, if possible, available,
- the problem is of reasonable complexity and has no satisfactory algorithmic solution.

The justification of an ES on purely economic grounds is not easy since the return on investment is usually difficult to evaluate. But other factors have also an importance :

- the expertise is scarce or fragile,
- the expertise is concentrated at one place but used at several locations,
- decisions must be taken under stressing conditions, ...

In the present state of AI methodology, the development of an ES is based on the interaction between two (or groups of) main actors :

- an expert of the application domain,
- a knowledge engineer.

These two actors will closely cooperate for building up the knowledge base of the ES. That constitutes the most crucial step in the development of an ES. It is likely that forthcoming tools from the research field of symbolic learning will facilitate the acquisition of knowledge from experts and its formalization {1}.

ES are software programs with a specific life cycle. The development of an ES is typically made up of four successive phases {8} :

- design of a first demonstrator : in this preliminary phase only a subproblem in the framework of the application is selected. The goal is only to demonstrate the validity of an ES solution and to evaluate the overall complexity and cost of the total project,
- development of a prototype able to take into account the entirety of the problem,
- integration of the final product in the environment in which it will be used. This phase raises complex problems related to classical computer science more than to AI : connection with other software packages and/or databases, interface with the users, real time aspects, etc.,
- exploitation of the ES : knowledge in a domain usually evolves when time goes, the maintenance and updating of the knowledge base is thus a vital operation in the life of an ES.

Development tools

The functionalities of an ES can be grouped into five categories :

- problem solving which represents the basic activity of the ES, carried out by one or several inference engines in relation with knowledge base(s),
- acquisition, modification and updating of knowledge,

- explanation on how a problem was solved by the ES. This function is fundamental if the ES is used in education, it is also very useful for the expert and the knowledge engineer to test the validity of the knowledge base during its design,

- connection of the ES with external systems : computer programs, data bases, eventually other ES,

- interface with the end user. This is an important point for a system to be accepted by the users. It includes various aspects : graphic displays, images, natural language processing, etc.

The complexity of an ES makes it often necessary to develop it by using sophisticated software tools rather than writing it from scratch in programming languages. There exists a large variety of such tools commercially available for different classes of machines : microcomputers, workstations, symbolic machines and mainframe computers.

Conclusion

The use of ES is rapidly increasing in a large variety of domains since they bring solutions to get unsolved problems. However these systems still present strong limitations :

- the learning abilities are modest (the knowledge base is built up a priori and cannot be dynamically updated by the ES),
- the size of the domain covered by the expertise of an ES is restricted,
- ES mainly use the surface knowledge of a domain instead of referring to the deep knowledge about underlying phenomena,
- reasoning schemes are too scarce and limited.

Some of these limitations are tending to be overcome in second generation ES which are now appearing. These systems are characterized by some new tendencies :

- multimodal representation of knowledge, especially in the framework of object oriented representations,
- use of deep knowledge in conjunction with a qualitative modelling of the phenomena,
- better integration of the ES in existing information systems,
- simultaneous use of several AI techniques : natural language processing, vision, planning, etc.

AI IN THE MOLECULAR WORLD

From what has been developed before, it is easy to imagine the numerous applications of AI techniques inside the molecular world {9}. We can try to classify them and to give some (non exhaustive) examples in such a way :

Programming languages and aids

Programming languages (object-oriented languages, LISP, PROLOG, etc.) allows to translate general and heuristic knowledge in an easier way than other procedural languages. Moreover, we can find shells designed for building systems entering specific categories.

Knowledge representation - Problem solving

As seen before, knowledge representation is present in every realization. The various possible modes help to make the best choice according to the problem.

Heuristic state-space search techniques may be used for synthesis planning programs : {10}, {11}, {12}, {13}, ..., structure elicitation : {14}, {15}, {4}, ...

Let us take the example of synthesis : a given state corresponds to a molecular structure while operators describe chemical transforms. To find a synthesis path is equivalent to find and reverse the "best" sequence of operators starting from the compound to be synthesized (initial state) and leading to the necessary basic chemical products.

Exhaustive search inside the resolution tree is impossible, so it is necessary to introduce heuristics (the technique of "best first search" with the computation of an evaluation function for measuring structural complexity is performed for example in SYNCHM {10}). In some implementations {11}, it has been decided to give an important role to the chemist using the system : an interaction mode allows to take into account the advice of the user to cut non productive branches and then to perform an automatic learning procedure to make the system better.

Learning

Automatic learning is relative to the automatic building by the system of some knowledge which has not been explicitly introduced. For example, training concerning biological sequences (training of filters isolating shorter codes assembly, training of concepts, etc.) {16}.

It is remarkable that the first developed ES, DENDRAL, which has been already mentioned, included an automatic learning module, MetaDENDRAL. To this module were presented some couples "problem-solution" (here, rough formula and mass spectrum on one side and the developed formula on the other side) and new reasoning spectrum interpretation rules have automatically been found, even, they say, concerning non elicited structures.

Learning may play a role in ES designed for data interpretation {17}, {18}, molecular synthesis {10}, {11} or search of links between structure and activity of a substance.

Structure elicitation

It concerns the determination of the composition and/or 3-dimensional structure of molecules. The first operational ES, DENDRAL {4} had been designed for finding the structure of chemical substances from their rough formula and their mass spectrum.

The strategy is a three-phase process :

- . a planning phase : structural constraints are inferred from available chemical and spectral data,
- . an algorithmic generation phase : all stereo-isomers compatible with the given constraints are generated,
- . a test phase : prediction of properties for each generated candidate (thanks to detailed structural/spectral relationships) and comparison of predicted properties and observed data to derive a plausibility score. And also simulations of experiments to differentiate among the remaining candidates.

AI techniques are applicable in the first and third phases.

A similar approach has been applied to DNA strands {19}.

Other systems have been developed for mass, NMR or laser spectrography {20}, {21}, the latter allowing to determine the composition of a mineral or organic substance from the study of its spectrum obtained by laser spectrography.

Structure-activity relationship

Some systems handle structural information, arising from molecular modelling, for product activity prediction. {22}, for example, includes a module for extracting a quantitative estimate of the activity of a molecule and a training module to refine these types of criterion each time a new molecule is studied.

Classification of molecules

In this domain, AI provides techniques for easy symbolic handling thanks to adequate languages, especially PROLOG, LISP and, more and more, the object-oriented languages, the prototype of which is Smalltalk. The last ones give a good support for complex object representation. And the introduction of ES, including problem specific heuristics decreases the complexity of the classification process. So, we can find a mixed knowledge representation mode, compound with both objects and production rules. Moreover, the classification process may involve approximate reasoning, by the way of theories like the ones of the possibilities {23} or of the fuzzy sets {24}.

A demonstrator aiming at classifying the photosynthesis inhibitors has been developed {25}. The objects are defined in terms of some structural elements

(presence and relative positions of functional groups and so on) and the reasoning rules deal with this topological information to predict the inhibitory properties of the molecule.

Chemical synthesis

We already mentioned some systems designed for chemical synthesis. They are direct applications of heuristic search. Various approaches {26}, {10}, {27}, {12}, {13}, {28}, {11} have been studied. Such systems generally includes multimodal representation of general chemical knowledge and heuristics.

Experiment planning

Such techniques may concern the planning of experiments in molecular biology : MOLGEN {6} has been the first ES including reasoning based on hierarchical planning, laying upon three knowledge levels : strategies, design and elementary actions. A mechanism for constraint managing acts for eliminating the inconsistent solutions.

Intelligent instrumentation and robotics

They allow the extension of the range of automated analytical procedures.

The evolution of VLSI integrated circuits leads to envisage expert systems to increase the efficiency of the interfaces with and the functioning of various types of instruments. The applications may be concerned with, for example :

- the introduction of expert systems acting directly when the information is captured (spectra interpretation, for instance),
- the aid for the processing and interpretation like the ones of electrophoresis gels {29}. This technique is presently considerably increasing together with the study and the production of proteins. The automatization of this technique would be an important step,
- the aid for protein ingeneering : crystallization or purification techniques in which the use of ES may intervene together with the automatization of the processes,
- the quality control : at the different steps of the production process (genetic control, microbiology, biochemistry, ...). One finds again the importance of image processing and of data interpretation associated to sophisticated techniques of reasoning.

Fault diagnosis in chemical analytical instrumentation

Dealing with troubleshooting of a complex installation shows that it needs much knowledge and much practice. Moreover, the number of failure causes may be very high.

Expert systems may give an answer to this problem by allowing the centralization and a redistribution of the knowledge, proposing a rigorous method for solving a problem and leading to a more efficient and rapid repair.

Information entry

Rule-based systems may be used for the syntactic control of data entered by the user of a graphic system, like the validity of molecules.

Oral language processing

Oral language may be in certain situations a convenient way of entering data, for example biological sequences like DNA strands for further automatic analysis. Common use of speech is still a domain of research.

Information retrieval and processing NL

These techniques allow information retrieval systems being used in a "friendly" way without a specified query language.

Molecular graphics

The graphic presentation on computer screens of the shape of active sites inside molecules make it possible a new molecular approach. Rule-based systems are able to give the optimal representation, to directly modify the molecule configuration (and then compute its energy), to superpose at best this molecule with other rigid or smooth modes, and so on. The results of the computational methods and of the visual observation may lead the specialist to a better understanding of some properties and to modelize this knowledge that can be included to enrich an ES dealing with the problem in question. The applications in the domain of the pharmacology, for example, are numerous, especially toward a better understanding of the central nervous system {30}, or aiming at modelling the reactivity of molecules (heterogeneous kinetics), etc. {31}.

Intelligent tutoring systems

The generalized declarative way of representing knowledge and the explanation abilities of ES allows to build systems design for "intelligent" computer-aided instruction (ICAI), and here, a chemical education {28}. These systems are generally compound with four parts (not always clearly separated) : a domain knowledge module, a pedagogical knowledge module, a student model or "profile" and the interface with the user.

REFERENCES

- {1} Y. Kodratoff, "Leçons d'apprentissage symbolique automatique", CEPADUES, 1986.
- {2} J. P. Haton et M. C. Haton, "L'intelligence artificielle", Que sais-je ?, PUF, à paraître.
- {3} A. Newell and H. A. Simon, "GPS : a program that simulates human thought", in E. A. Feigenbaum and J. A. Feldmann, eds, *Computer and thought*, New York: Mc Graw Hill, 1963.
- {4} R.K. Lindsay et al., *Applications of Artificial Intelligence for Organic Chemistry : the DENDRAL Project*, Mc Graw Hill, New York, 1980.
- {5} E. H. Shortliffe, "Computer-Based Medical Consultations : MYCIN", ELSEVIER, 1976.
- {6} M. Stefik et al., "Knowledge programming in LOOPS", *AI Magazine*, fall 1983.
- {7} R. O. Duda et al., "Model design in the PROSPECTOR consultant system for mineral exploration", in D. Michie, ed., *Expert systems in the micro-electronic age*, Edinburgh University Press, 1979.
- {8} A. Bonnet, J. P. Haton et J. M. Truong-Ngoc, "Systèmes Experts : Vers la Maîtrise Technique", InterEditions, 1986.
- {9} N. A. B. Gray, "Artificial Intelligence in Chemistry", *Analytica Chimica Acta*, 210, 9-32, 1988, in Elsevier Science Publishers, Amsterdam.
- {10} H. Gelernter et al., 1st Conf. Artificial Intelligence Applications, IEEE, Silver Springs, MD, 1985.
- {11} M. C. Haton et al., "Représentation des connaissances et apprentissage en synthèse organique : une démarche experte", *Actes du Colloque Cognitive*, 1987.
- {12} W.T. Wipke and D. Rogers, "Artificial Intelligence in Organic Synthesis. SST : Starting Material Selection Strategies, An Application of Superstructure Search", *J. Chem. Inf. Comput. Sci.*, 1984.
- {13} J. B. Hendrickson and A. G. Toczek, "Synthesis design logic and the SYNGEN (synthesis generation) program", *Pure & Appl. Chem.*, vol. 60, no. 11, 1988.
- {14} M. E. Munk et al., *Anal. Chim. Acta*, 184, 1986.
- {15} Y. Kudo and S. Sasaki, "Principle for Exhaustive Enumeration of Unique Structures Consistent with Structural Information", *J. Chem. Inf. Comput. Sci.*, 25, 1976.
- {16} J. Sallantin et al., "Biostation, un poste de travail dédié à l'analyse de séquences génétiques", *Rapport CEE BAP, CRIM, Montpellier*, 1987.
- {17} T.M. Mitchell and G.M. Schwenzer, *Org. Magn. Reson.*, 11, 1978.
- {18} B.G. Buchanan et al., *J. Am. Chem. Soc.*, 98, 1976.
- {19} M. Stefik, "Inferring DNA structures from segmentation data", *Artificial Intelligence*, 11, 1978.
- {20} A. Terry, "The CRYSLIS project : hierarchical control of production systems", *Ph. D. Thesis, Stanford Univ.*, 1983.
- {21} J. P. Haton et al., "EXSYLA, un système expert pour l'interprétation de spectres de masse en spectrométrie laser", *Journées Int. Systèmes Experts*, Avignon, Mai 1985.
- {22} R. Rozot et al., "Handling the structural information arising from molecular modelling for drug activity predictions : the SARAH system", 44th Int. Meeting of the SFC, Nancy, Septembre 1989.
- {23} D. Dubois et H. Prade, *Théorie des possibilités*, Masson ed., 1985.
- {24} L. A. Zadeh, "Commonsense knowledge representation based on fuzzy logic", *Computer*, 16, 1983.

- {25} M. Gingold, "Classification de molécules chimiques : présentation d'une maquette de système expert", 44ème réunion internationale de la division de Chimie Physique de la SFC, Nancy, Septembre 1989.
- {26} K. Funatsu and S.I. Sasaki, "Computer-Assisted Organic Synthesis Design and Reaction Prediction System, "AIPHOS"", Tetrahedron Comp. Methodology, vol. 1, n. 1, 1988.
- {27} R. Herges, "Reaction Planning : Computer-Aided Reaction Design", Tetrahedron Comp. Methodology, vol. 1, n. 1, 1988.
- {28} E.J. Corey et al., Science, 228, 1985.
- {29} J. P. Haton, et P. Nugues, "Un système multi-expert pour l'interprétation d'images d'électrophorèse 2-D", Actes du Congrès AFCET-RFIA, Paris, Novembre 1989.
- {30} R.C. Maroun, "Study on the sequence selectivity of the bi-functional intercalator ditercalinium", 44th Int. Meeting of the SFC, Nancy, Septembre 1989.
- {31} P. Azario et al., "The MARSEIL/SOS Expert System, a New Graphic Approach", J. Org. Chem., 1988.

DISCUSSION

SOLIE - Vous avez dit au cours de votre conférence que les capacités d'apprentissage des systèmes-experts étaient modestes. Pouvez-vous expliquer et commenter ce point par des exemples sur ces systèmes-experts se rapportant au domaine de la chimie ?

HATTON - Presently, expert systems can increase their own ability to problem solving only by simple learning mechanisms, generally involving logical reasoning. It must be noticed that the first expert system, DENDRAL, designed for structure elicitation, included a learning module, named METADENDRAL. From a collection of examples (data = rough formula and mass spectrogram of an organic molecule, solution = developed formula), it has been able to infer new rules for the interpretation of mass spectrograms.?

Automatic learning is still a research topic. However, some applications are appearing in some domains in which the use of automatic systems implies a strong interaction with the user who confirms or rejects the solution proposed by the system (synthesis of organic molecules, for instance).

DEVILLERS - My question is somewhat related to theology. Do you need in expert-systems to explicitly introduce non-contradiction principle (or rule) or is it self-contained ?

HATON - When an expert system is being used, reasoning rules (especially logical ones) are used to infer new information concerning the problem in question and to draw conclusions, make a decision, etc.

The consistency of the knowledge base must be warranted at the building of the system. Actually, it is known that "from wrong premises, anything may be inferred".

There presently exists some tools to help the knowledge engineer to detect a possible inconsistency within the knowledge base.

AN AID TO THE CLASSIFICATION OF MOLECULE: PRESENTATION OF AN
EXPERT SYSTEM PROJECT

M.P. GINGOLD

Service de Biophysique, CEA/CEN-Saclay, F 91191 Gif/yvette
Cedex (France)

SUMMARY

As an initiation to the methods of building an expert system, a program which studies molecules depicted as molecular graphs is presented. The expert systems development tool (SPIRAL) encompasses the PROLOG language, with an object formalism. The molecules are described as objects with two main slots, their labelled atoms and the types of bonds between them.

INTRODUCTION

More and more effort is being made to predict physical, chemical and biological properties of molecules, as exemplified by the development of quantitative structure-activity relationship (QSAR) studies (ref. 1). The problem arises then to manipulate and extract molecular structures and information contained in data bases. We focus here on the topology determined by the description of the graph of a molecule.

We demonstrate how an declarative language can be used to give information pertaining to such QSAR studies: detection of given substructures, enumeration of subgraphs and associated calculations such as molecular connectivity index determinations. The listing of the examples can be obtained from the author upon request.

LOGIC PROGRAMMING LANGUAGE

The expert systems development tool used was SPIRAL. This tool was developed by Dr. Yves SOUCHET (CEA, Commissariat à l'Energie Atomique, Service de Mathématiques Appliquées). Programmed in C, it encompasses the PROLOG declarative language with improvements (first-order logic formalism) and, in addition, supplies an object formalism (with daemons and consistency checking). Thus it is possible to define prototypes (classes) and class instances (objects). The classes have

system- and user-defined slots with simple values inheritance. A version is adapted to the PC compatible micro-computers, whereas the full version is mainly developed for UNIX workstations (with graphics, input and output primitives). Procedural subroutines can be called from within the program.

PROGRAM

Description of the molecular structures.

A molecular structure is defined as an object. It can be described directly by the user when running the program, as objects, facts and inference rules can be added (or subtracted) interactively. This is obviously tedious, and it is preferable to prepare an editable data base which can be further structured and, if necessary, partitioned, to avoid memory size limitations.

Basically, a molecule is described by the sequence of its atoms, their labelling (to distinguish identical atoms) and the list of the type of the bonds between them (Figs. 1,2). The labels of the atoms are chosen by the user. This corresponds to the vertices and the edges of the graph representing a molecular structure (refs. 2-4). These two main slots, atoms and bonds, are lists of lists. The elements of the first list contain an atom name and its label. The elements of the second list contain a bond type and the two labels corresponding to the bonded atoms.

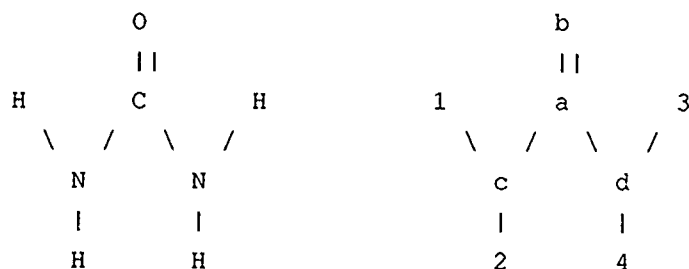


Fig. 1. Example of urea. Scheme of the molecule.

```

def_classe(molecule, objet, (atoms, liste), (bonds, liste);
def_instance(urea, molecule
, (atoms, (C, a), (O, b), (N, c), (N, d), (H, 1), (H, 2), (H, 3), (H, 4))
, (bonds, (D, a, b), (S, a, c), (S, a, d)
, (S, c, 1), (S, c, 2), (S, d, 3), (S, d, 4)))

```

Fig. 2. Instance definition for the urea molecule. S: single; D: double bonds.

Additional, optional slots may contain the name(s) of the file(s) for the bibliographic reference list, the physical data, the scheme of the molecule which can be displayed on the screen, the screen coordinates of the atoms in order to highlight some particular atoms following a query, and so on. Defined atoms can be functional groups or even molecular substructures, depending on the chosen level of description. Groups and substructures can in turn be defined as objects: the description of the molecule is then shortened and more user-readable, without losing its full content if needed by later treatments. Hydrogen atoms can be suppressed at will, as retrieving them is straightforward, as the bond orders and the connected atoms are known. Aliases can be given to a molecule or a group name.

Single molecule queries.

Logic programming easily allows inter alia to find out the shortest distance between atoms (in terms of number of bonds), the presence of cycles, the molecular connectivity indexes (refs. 5-7) after having enumerated the appropriate subgraphs (paths, clusters,...). The queries can be logically mixed together. They are derived from basic inference rules which manipulate the slots of the molecule object. Due to the fundamental properties (backtracking) of the declarative language, the knowledge base can be automatically scanned.

Two molecule queries.

As noted above, each molecule has its own atom labels. Consequently, in order to be able to directly compare two

molecules, these labels are substituted by numbers. The sequence numbers can be manipulated by permutating, extracting, sorting, etc. Moreover, identical answers are usually automatically eliminated. Thus particular substructures, or molecules embedded in others may be located.

CONCLUSION

Applied to molecular graphs, the use of a logic programming language renders enumeration tasks particularly suitable. In a given program one can easily extend both the molecular data base and the set of queries it is able to answer, with the goal of an artificial intelligence approach to structure-activity studies.

REFERENCES

- 1 G. Naray-Szabo, The harmony of molecules, in: J. Maruani (Ed.), *Molecules in Physics, Chemistry and Biology. General Introduction to Molecular Sciences*, Vol. 1, Kluwer, New York, 1988, pp. 205-231.
- 2 S.C. Basak, V.R. Magnuson, G.J. Niemi, R.R. Regal, Determining structural similarity of chemicals using graph-theoretic indices, *Discrete Applied Mathematics*, 19, (1988) 17-44, in: J.W. Kennedy and L.C. Quintas (Eds.), *Applications of Graphs in Chemistry and Physics*, North-Holland, Amsterdam, 1980.
- 3 H.J. Luinge, G.J. Kleywegt, H.A. Van't Klooster and J.H. van der Maas, Artificial intelligence used for the interpretation of combined spectral data. 3. Automated generation of interpretation rules for infrared spectral data, *J. Chem. Inf. Comput. Sci.*, 27 (1987) 95-99.
- 4 Y. Sun, L. Pierrons and M.-C. Haton, Résolution de problème et raisonnement expert en synthèse de molécules organiques: le système OASIS, *Journées Internationales Systèmes Experts et Applications*, Avignon (1988) pp. 483-495.
- 5 K. Takeuchi, C. Kuroda and M. Ishida, Prolog program for subgraph enumeration and calculation of molecular connectivity indexes, *J. Comput. Chem*, 10, (1989) 380-385.
- 6 G. Klopman, C. Raychaudhury and R.V. Henderson, A new approach to structure-activity using distance information content of graph vertices, *Mathl Comput. Modelling*, 11 (1988) 635-640.
- 7 M. Randic, S.C. Grossman, B. Jerman-Blazic, D.H. Rouvray and S. El-Basil, Modelling Drug Design II, *Mathl Comput. Modelling*, 11 (1988) 837-842.

HANDLING THE STRUCTURAL INFORMATION RESULTING FROM MOLECULAR MODELLING FOR DRUG ACTIVITY PREDICTIONS: THE SARAH SYSTEM

Roger ROZOT, Jean-Louis RIVAIL and Hervé MATHIS

Laboratoire de Chimie Théorique, UA CNRS n° 510, Université de Nancy I, Domaine
Scientifique Victor Grignard, BP 239, 54506 Vandœuvre lès Nancy Cedex (France).

SUMMARY

SARAH, a new program which combines artificial intelligence (pattern recognition and machine learning) and statistics has been elaborated in order to assist the chemist in molecular modelling and the search for structure-activity relationships. It extracts the relevant information from data of a training set to create both qualitative and quantitative rules which appear to govern the activity of parent molecules.

INTRODUCTION

The SARAH software, for "Structure - Activity: Relationships by Apprenticeship and Heuristics" [1,2] is presented. Its purpose is to assist the chemist in devising molecules which have properties interesting in pharmacology. After a molecular modelling step, one has three different levels of structural information on a molecule:

- (i) the nature and the connectivity of the atoms, which comes from the input data,
- (ii) the geometry of one or several conformations,
- (iii) the electronic properties, especially if the modelling is based upon a quantum chemical computation.

This information is very rich and is assumed to contain most of the data which are liable to play a rôle in the biological activity. The problems which we have to solve are:

- (i) How to extract the relevant information.
- (ii) How to correlate it to the activity.

The SARAH approach assumes that there exists a *training set* of molecules the activity of which has been measured for the biological process of interest. Furthermore, one assumes that for each molecule one knows the geometry of the most stable conformation(s) as well as the electronic wavefunction usually obtained by a semi-empirical quantum chemical package such as GEOMO [3]. These data are used to initiate a *learning process* which should discriminate between active or inactive molecules on a basis of purely structural characteristics. The *quantitative relationships*

between the score of the activity test and the structural descriptors is established, within each class by means of a classical statistical method. This initialization procedure being achieved, it becomes possible to submit a new molecular structure and the system locates it in one of the classe and gives its position relative to the other molecules of the class.

THE LEARNING PROCESS

This step is a generalization of the method used by a chemist who wants to compare different molecular structures in a series of pharmacologically related molecules.

It starts by finding a sub-structure, or *reference pattern*, which is common to all the molecules of the series in order to allow an unambiguous superimposition of the molecular geometries. This can be done automatically by two dimensional pattern recognition if the reference pattern has the same connectivity in all the molecules of the training set. When the reference pattern is defined by a limited number of atoms (usually heteroatoms) which occupy the same relative positions in space in all the active conformations but have a different connectivity, the previous procedure fails. In this case the reference pattern may be defined by the user through an interactive graphic routine which displays the superimposition of two molecules.

Then the user has to designate a *reference molecule* (usually in the active subset) and all the molecules of the training set are compared to it by scanning the space around origin chosen on the reference pattern. Each direction has two intersections with the van der Waals molecular surface where several structural descriptors can be defined: distance to the origin, electrostatic potential, electric field ...

The similarities and the differences between a molecule of the training set and the reference molecule are pointed out. This makes possible a machine learning of the activity concept by search for the structural descriptors which discriminate the two classes of molecules that is to say by determination of the *activity rules* [4-10].

Finally, in each subset, a multivariate regression between the value of the activity score and a series of other structural descriptors which may play a rôle in modifying the activity is established.

The flow chart of the learning process is given on figure 1.

SUBMITTING A NEW STRUCTURE

The activity rules are usually very useful to devise an improved molecular structure. After the molecular modelling process it is submitted to the activity rules and an expected activity score is given (see figure 2).

When, in the new molecule, a molecular descriptor lies outside the range of variation observed on the training set, the system automatically indicates that its predictions may be wrong, due to an effect which has not been taken into account in the learning process. This is particularly true in the case of steric effects for which the biological response does not vary monotonously with the molecular size.

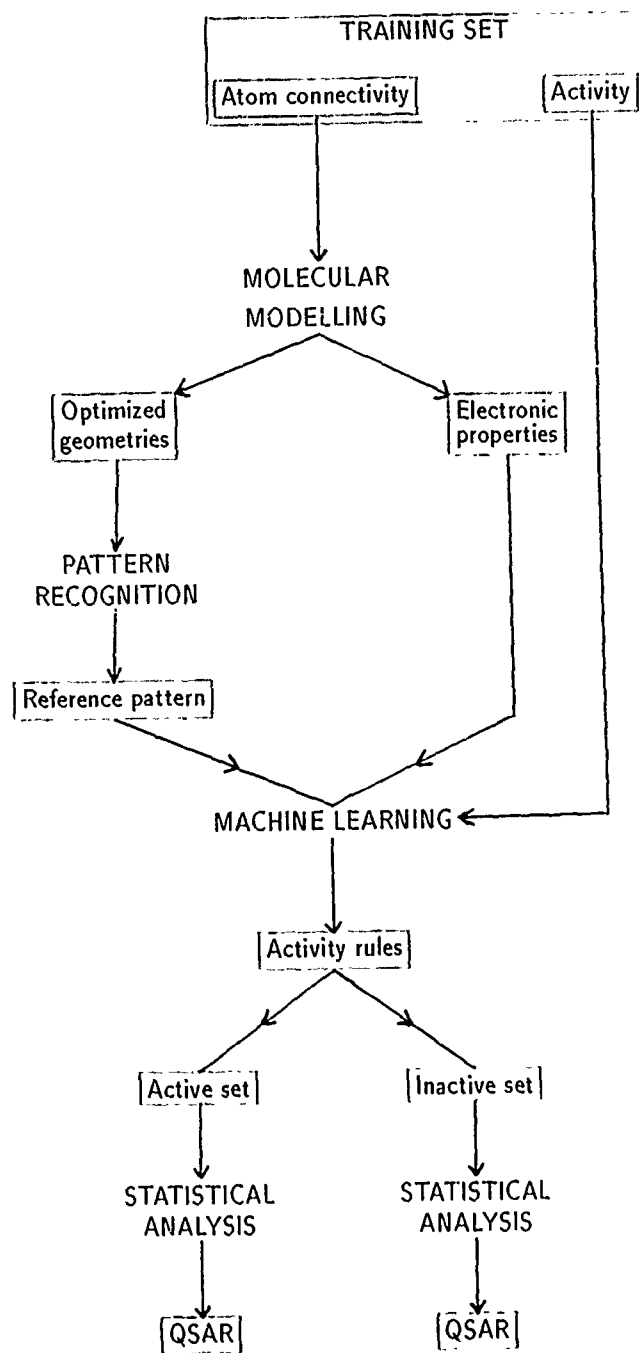


Figure 1: The learning step of the SARAH system

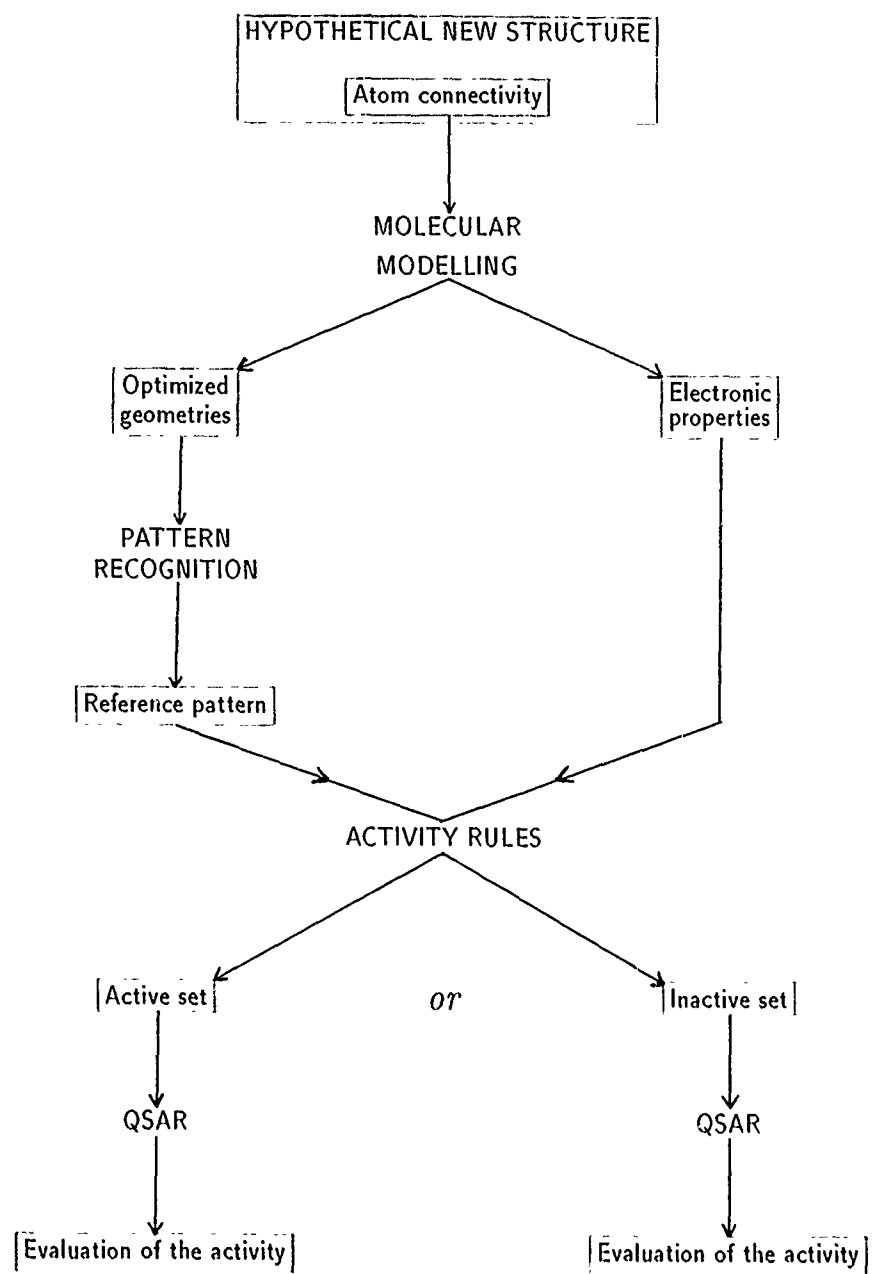


Figure 2: Submitting a new structure

UPDATING THE SYSTEM

When a structure which does not belong to the training set appears to have its activity quantitatively determined, it is possible to add it to the training set and to update both the activity rules and the QSAR.

EXAMPLE OF APPLICATION

The first attempt to use the SARAH system was for the antiepileptic family of 1,4-benzodiazepines. The training set contains 49 molecules, 34 in the active subset and 15 in the inactive one. *Diazepam* (or 7 chloro - 1 methyl - 5 phenyl - 3 H - 1,4 benzodiazepin 2 - one) has been chosen as reference molecule in the active set.

The learning step automatically generates a very simple activity rule only based upon geometric data related to the van der Waals surface of the molecules:

If the length of the substituent of the carbon 7 (D_7) is between 3.20 Å and 5.28 Å and if the length of the substituent of carbon 5 (D_5) is smaller than 6.58 Å then the molecule is active.

For these molecules, the activity is lineary correlated to four electronic descriptors which are the electrostatic potentials in the vicinity of each nitrogen atom of the diazepine ring and in the vicinity of atom 3' of the phenyl ring and finally the electric field in the vicinity of atom 8 on the benzo ring (see figure 3).

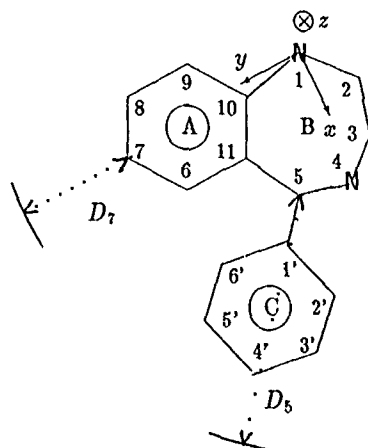


Figure 3: Result of the learning step:
 $(3.20 \text{ Å} < D_7 < 5.28 \text{ Å}) \text{ and } (D_5 < 6.85 \text{ Å}) \Rightarrow \text{active molecule}$

These descriptors are easy to interpret. The first two are clearly related to the basicity of the nitrogen atoms (or to their ability to create a H bond). The last one is clearly connected to the polarity of the substituent in 7. The third one may be a measure of the electronic influence of the substitution of the phenyl ring in 5.

The learning step has been followed by the analysis of a test set of 12 new molecules. 10 of them are classified in the appropriate subset and the estimated activity is in fair agreement with experiment. The last two molecules, which are considered as inactive and are classified erroneously are the only ones which contain a large substituant (ethyl and phenyl) on carbon 3 although the largest substituant of the molecules of the training set is a methyl group. As indicated previously, this particularity is mentioned by the system. This is an indication of another steric effect which would probably produce a complementary activity rule but this assumption has not been tested yet.

CONCLUSION

The SARAH system represents a progress in the search for correlations between structure and activity.

By separating the analysis into two steps: the search for general activity rules and the quantitative correlation of this activity with structural data, it reproduces rather closely the intellectual process of the chemist. This procedure takes into account the differences between steric and electronic effects in biological activity.

It seems plausible that within a reasonable range this activity varies monotonously with the electronic effects so that the corresponding descriptors are well adapted to QSAR. Conversely, steric effects often imply that along some directions the molecule must be neither too long nor too short. This feature can easily be expressed by activity rules.

This difference clearly appears in the 1,4-benzodiazepine family in which the rule is purely steric and the factors of QSAR are all electronic.

This example illustrates rather well the superiority of the SARAH approach on the purely statistical methods [1,11].

REFERENCES

- [1] ROZOT, R., "Facteurs Electroniques et reconnaissance de formes en C.A.O. de Molécules Pharmacologiquement Actives. Application à la Famille des Benzodiazépines.", mémoire de Doctorat, Université de Nancy I, 1988
- [2] ROZOT, R. and RIVAIL, J.-L., "From Machine Learning to Drug Design: the SARAH System.", submitted to *Journal of Computer Aided Molecular Design*, 1989
- [3] RINALDI, D., QCPE 290
- [4] QUINLAN, J. R., *Machine Learning*, 1, MICHALSKI, R. S., CARBONELL, J. G. and MITCHELL, T. M. (Eds), Tioga Publishing Company, Palo Alto, California (1986), 81-106
- [5] QUINLAN, J. R., *Machine Learning*, 2, MICHALSKI, R. S., CARBONELL, J. G. and MITCHELL, T. M. (Eds), Tioga Publishing Company, Palo Alto, California (1986)
- [6] GASCUEL, O., in *Congrès AFCET - Reconnaissance de formes et Intelligence Artificielle*, Grenoble, France (1985), 1263-1271

- [7] GASCUEL, O., in *Actes des Premières Journées Françaises de l'Apprentissage*, L.R.I., Orsay, France (1986)
- [8] GASCUEL, O., in *Proceedings of the first European Working Session on Learning*, Orsay, France (1986)
- [9] GASCUEL, O. and DANCHIN, A., *Journal of Molecular Evolution*, 24 (1986), 130-142
- [10] GASCUEL, O. and DANCHIN, A., *Eighth European Conference on Artificial Intelligence*, Munich, Pitman, London (1988), 390-395
- [11] ROZOT, R. and CARTIER, A., "The Family of Diazepines: a Detailed Study by the Learning System SARAH.", to be published

OCTANG SYSTEM: A NEW APPROACH OF 3D MOLECULAR GEOMETRY.

E. DELANCE , J.P DOUCET , J.E. DUBOIS
ITODYS, Associé au CNRS, UNIVERSITE PARIS 7.
1 Rue Guy de la Brosse 75005 PARIS FRANCE

ABSTRACT:

For a few years now, molecular shape has emerged as one of the main parameters in simulation strategies. This is a general agreement that 3D geometry is of prime importance for molecular recognition, and many structure-property relationships. Despite these efforts, automatic identification of tridimensionnal structural fragments is still an open problem. Most of the time, one cannot compare shape similarities in a quantitative manner.

As a new description tool, we propose the OCTANG method. This system deals with topological information (graph), where 3D information lies in valuated edges. The OCTANG system has already succeeded in the identification of a predefined molecular fragment, among a set of compounds. It should allow us for developing new algorithms for structural database management, pharmacophore identification, and maximal common substructure searching.

INTRODUCTION

Infography has emerged as an effective tool in chemistry. One can now describe and simulate some of the quantum mechanical concepts, such as electron densities or electrostatic potential (1). These electronic properties are essential in "molecular recognition" mechanisms, where the complementarity of the surfaces may optimize the interactions.

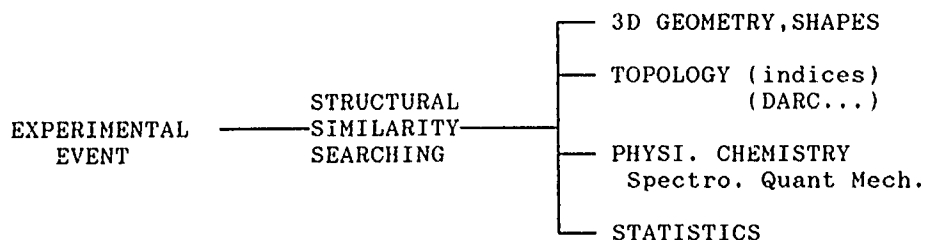
The technological supplies have revealed how important tridimensional geometry and shape are for molecular description. It also emphasized one of the main problems : automatic recognition of 3D molecular objects.

This is an important problem because knowledge in chemistry has to be widely distributed, for general use such as Drug Design, Spectroscopy, Statistics, etc. For this purpose, one often needs to get a global information, from structural and sometimes substructural information. So the representation of this structure should be treated as one of the main parameter, for the efficiency of structure property relationships and databases management systems.

However, the actual nomenclature does not involve 3D geometry. This may be the reason why one is not accustomed to it, and it could explain why molecular geometry is only described in a unique manner, that is cartesian coordinates. Consequently, we propose a new description method, called the OCTANG method. The molecular descriptor that we create, looks like a topological one, but the edges of the chemical graph are 3D significant. This allows for an easy and fast handling of molecular information, and we suggest in this paper a set of applications. Before presenting OCTANG method, we are going to summarise briefly the current concepts which are used in automatic structural recognition.

Molecular recognition and similarity:

Tridimensional structural or substructural recognition takes place in information retrieval from databases, but also in structure-property relationships. In both fields, the common concept remains similarity searching, according to the key-lock concept. In fact, we always try to explain an experimental event, with structural characteristics :



Similarity searching often involves topological indices. They are easily calculated, and closely related to the chemical graph (2). Basically, they are based on node and pathway count. One of the first was the WIENER parameter, designed to predict boiling points. (3)

Randic, one of the main workers in this area, proposed a first topological indice (4), then a second one, called "molecular

identification number" (5). The latter was supposed to give a unique description of molecular graph, but a recent systematic calculation for alkane series up to 20 carbons atoms, revealed a hundred of non isomorphic pairs with identical indices. (6)

Anyway, the concept of pathway count is now well known as molecular connectivity, and remains one of the most widely used methods in correlation analysis. (7)

The most critical problem arises when two different graphs have the same indice values. Obviously, topological indices are not unique, and the best way to use them in databases remains to design screen strategies. Database management with 2D descriptors is a reality since ten years now, with DARC system or CAS system (8,9).

Another way for similarity searching, is using 3D molecular geometry, that is atomic cartesian coordinates. In this field, we have to separate Maximal Common Substructure searching, and Predefined Common Substructure searching.

The first one, MCS, consists in identifying the largest common structural feature in a set of compounds. So, combinatorial problems arise, when generating intermediate solutions during the so-called growing steps. In the second case, a molecular fragment has to be searched among largest structures. Here are some of the current methods:

Crandell et Smith	(10)	Ullman	(14)
Clique detection	(11,12)	Set reduction	(14)
Lesk	(13)	Stouch/Jurs	(15)

All these methods use interatomic distances, represented in a distance matrix, (excepted Stouch/Jurs' one). Consequently, the common purpose for these algorithms is to find an isomorphism or a subgraph in a larger graph. Each of them generally needs a dissuasive CPU time, so these approaches cannot be easily performed.

The Stouch method is based on a 3D grid, to define and compare molecular shapes in a discrete manner.

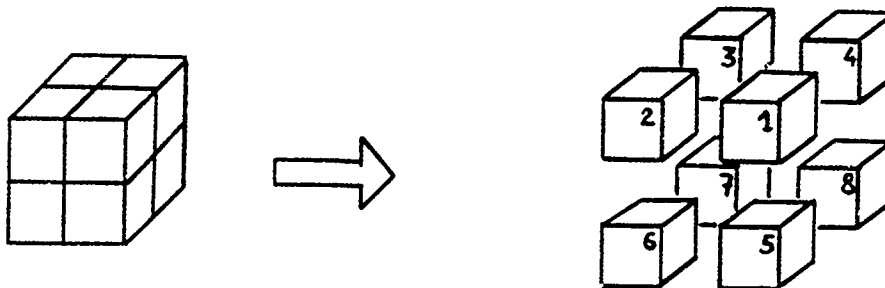
Finally, two distinct strategies are emerging:

- Structural databases management which needs rapidity. The only way seems to be bit screens methods, associated with more precise methods operating on relevant database subsets.
- Structural or electronic similarity searching which proceeds in a quite different manner. 3D Grid methods and shape superposition are used now, but a new kind of similarity indexes, based on electronic densities have been proposed.

We attempted here to design a new molecular descriptor, which is basically a topological one, including an edge chromatism to describe tridimensional relative positions. This is the OCTANG method.

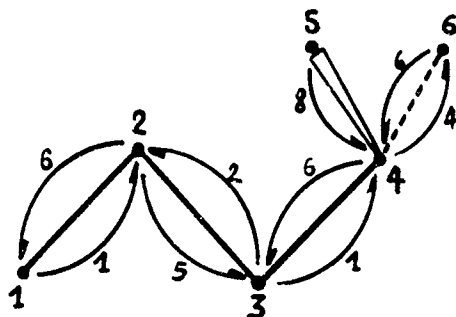
THE OCTANG METHOD:

The description of a set of points may be performed by their relative positions, instead of cartesian coordinates. It is a local perception of a point environment, which is easily extended in a 3D space. This space is then subdivided into 8 zones according to X,Y and Z axes:



Such a system can be moved on each atomic position in a skeleton, to describe its neighbours without any dimensionnal parameter. We chose to describe only the first topological neighbours for each atom, and to keep a complete connectivity array. This point is important to be noted, because the direction code depends on the view point.

With these conventions, we can fully describe a molecule with two arrays (connectivity and direction), and a chromatic list:



6, 6, 6,6, 6, 6

1	2	1	1
2	1 3	2	6 5
3	2 4	3	2 1
4	3 5 6	4	6 2 4
5	4	5	8
6	4	6	6

chromatic list
connectivity array
direction array

Structural identification:

The first problem arising when trying to compare two molecules, is their relative positions. With our codification system, a unique fragment is perceived in two different ways, if it lies in two different positions.

So it is clear that molecular descriptors cannot be strictly compared. For the recognition step, we developed a pathway representation associated with the "elementary fragment" concept.

An elementary fragment is a triple of nodes from the graph. It is the smallest object characterised by an internal geometry, and it is especially the basic descriptor which will allow for the pathway recognition.

We have also developed the equivalent elementary fragment concept, to gather in a same group all the elementary fragments with a same internal geometry. Thus, we can recognise any fragment, whatever its global orientation.

The pathway decomposition is led by a small set of rules, and finally any graph is perceived as a main pathway, with a set of ramifications.

The main pathway is the topologically longest one, identified by a deep-first search algorithm, from high connectivity nodes. The ramifications may be designed in a "recursive" manner. We mean that any ramification may be attached either on the main pathway or on another ramification. Each pathway or ramification is then totally described by:

- Length
- Chromatic node list
- Cyclisation node (ring closure)
- Parent pathway
- Anchoring node on the parent pathway
- Contact node (closure on the parent pathway)

It must be kept in mind that the pathway representation exists only during the recognition step. In practice, a molecule is stored as a triple : chromatic list, connectivity array, direction array.

The pathway decomposition and the elementary fragment concept finally lead to a complete description system. We present now, a schematic view of the recognition step.

A 3D request is presented as a set of related pathways, and the most straightforward strategy is to sequentially search a set of equivalent pathways in a larger structure.

This is performed in an iterative way, where each pathway is progressively recognized by its elementary fragments. Here, the equivalent fragment concept is of prime importance to decide whether two triples are congruent or not. During the step of progression for each pathway, we have chosen a partial overlapping to keep consistency along the whole pathway.

Finally, a pathway is accepted only if all its parameters match the reference parameters (ring closure, contact node...).

In practice, these parameters are used to direct search procedure towards the best ways. Before we conclude, we can summarise the recognition strategy as follows:

- Given a molecule described by chromatic list, connectivity array and direction array.
- Given a requested structural fragment, described by its pathways.

Does the fragment lie in the molecule ?

The general procedure to answer this question is the following:

- Select in the molecule, all the nodes with the same chromatism and connectivity as the requested main pathway first node.
 - With each of these nodes, search for a pathway similar to the main pathway requested
 - For each first node which has succeeded in generating a main pathway, apply the same procedure to find out the ramifications.
- This step is quickly achieved since we know the row of anchoring nodes in the main pathway.

This strategy has been successfully applied in 3D structural recognition in many different cases. But we have to mention here some distinctive features:

- One can obtain many solutions separated only by the utilisation of the nodes (pathway or ramification). In fact these solutions have to be filtered.
- Another redundancy may arise because many nodes can be used as first node of a pathway. (1-2-3 , 3-2-1)
- Partial solution is possible. Our actual algorithm accepts a solution if the main pathway is shorter than the requested one but takes care of the anchoring nodes.

In practice, we avoid redundancy by filtering multiple solutions just before storing them.

CONCLUDING REMARKS:

The first application of OCTANG method was the recognition of a predefined 3D structural fragment. The system has proved itself efficient. We are now working on the MCS problem, and we are aiming at a breadth-first search procedure using elementary fragments. This part of the system should allow for the determination of a structural pharmacophore.

We are also studying the stereochemistry codification. The equivalent fragments concept is not well suited for dealing correctly with stereochemistry. So we plan to include this knowledge in a screen system, designed as a node and its first order topological environment for selection in a database.

Last but not least, we are currently working on a database structure using our representation mode (chromatism list, connectivity and direction arrays). In such a structure it would be easy to design efficient screens and to manage structural and substructural requests.

References:

- 1 Dubois,J.E , Yue,S.Y , Doucet,J.P
The Visual Computer 1985,2,367
- 2 Balaban,A.T J.Chem.Inf.Comput.Sci 1985,25,334-343
- 3 Wiener,H , J. Chem. Soc 1947,69,17-20
- 4 Randic,M J.Am.Chem.Soc 1975,97,6609-15
- 5 Randic,M J.Chem.Inf.Comput.Sci 1984,24,164-175
- 6 Szymanski,K , Müller,W.R , Knop,J.V , Trinajstić,N
J.Chem.Inf.Comput.Sci 1985,25,413-415
- 7 Kier,L.B , Hall,L.H , "Molecular Connectivity in
Structure activity analysis" WILEY, New-York 1986
- 8 Dubois,J.E , Panaye.A , Attias,R
J.Chem.Inf.Comput.Sci 1987,27,74-82
- 9 Attias,R J.Chem.Inf.Comput.Sci 1983,23,102-108
- 10 Crandell,C.W , Smith,D.H
J.Chem.Inf.Comput.Sci 1983,23,186-197
- 11 Barrow,H.G , Burstal,R.M
Inf.Proc.Lett 1976,4,83-84
- 12 Brint,A.T , Willett,P J.Chem.Inf.Comput.Sci
1987 27,152-158
- 13 Lesk,A.M Communications of the ACM 1979,22,219-224
- 14 Brint,A.T, Willett,P J.Mol.Graphics 1987,5,49-56
- 15 Stouch,T.R , Jurs,P.C J.Chem.Inf.Comput.Sci
1986,26,4-12

MOLECULAR GRAPHICS FOR THE MACINTOSH

J.M. CENSE

Laboratoire d'informatique chimique, Ecole Nationale Supérieure de Chimie de
Paris, 11, rue Pierre et Marie Curie - 75231 Paris Cedex 05 - FRANCE

SUMMARY

Two programs, MolDraw and MolView, implemented on the Apple Macintosh, are described. These programs give access to good quality 3-D pictures of atomic and molecular objects whose coordinates have been read in from files formatted according to several standards. Depth-cueing, stereo views and animation are used to enhance perception of depth. High quality hard copies of all pictures are readily obtained by using laser printers.

INTRODUCTION

Now that Apple Macintosh and high resolution laser printers are widely available it appears interesting to have an easy access to specifically tailored packages allowing creation and manipulation of chemical drawings. Using these programs, most of the visualization job, classically performed on expensive workstations, can be accomplished on a cheap micro-computer. Furthermore the drawings so obtained may be directly printed on a wide variety of laser printers or put on the Macintosh clipboard to be later processed by word processors, desktop publishers, presentation managers or graphics software.

HARDWARE AND SOFTWARE

MolDraw runs on any Macintosh with 1 Mb of memory and is mainly black and white oriented. It accepts structures up to 512 atoms or 600 bonds. MolView is written to run only on the Macintosh II with a minimum of 2 Mb of memory and 256 colors (or at least 256 gray tones). It handles structures up to 8000 atoms or bonds.

With a palette of 256 colors, objects can be displayed in 8 colors with 32 levels of shading, this is sufficient to give an excellent rendering of shaded spheres. The visible region of spheres is determined quickly by using regions manipulation techniques (ref. 1), shading is performed by stamping computed templates (one for each atom type) into the calculated visible region.

Due to the fast bit-transfer tools implemented in the Macintosh Toolbox, a color space-filling picture of a 500 atom molecule can be obtained in less than 15 seconds at

the screen resolution of the Macintosh II (640 x 480, 72 pixels per inch). Most models (<50 atoms) are displayed in less than 3 seconds.

Both programs are written in Lightspeed C and follow the general Macintosh interface with windows, scrolling menus, dialogs, keyboard equivalents and desk accessories.

Several printers have been tested : Apple ImageWriter II and LaserWriter, QMS ColorScript 100 and Tektronix 4693 DX color laser printers. Full page black and white images at maximum resolution (300 ppi) are printed in less than 3 minutes.

3-D STRUCTURE INPUT

MolDraw and MolView programs read text files of atomic coordinates according to several formats : SYBYL (ref. 2), PDB (ref. 3) or CHEM3-D (ref. 4). These text files are transferred from host computer or created on the Macintosh using molecular builder such as CHEM3-D (ref. 4) or PCMODEL (ref. 5).

A private format for Cartesian, internal or fractional coordinates is also provided. Internal coordinates files permit rapid construction of small molecules whereas fractional coordinates files are used to collect crystallographic information.

These crystallographic files include cell constants, fractional coordinates, symmetry operators and repetition operators to generate several contiguous cells. Therefore, processed structures are not restricted to molecular structures.

3-D STRUCTURE MANIPULATION

MolDraw and MolView programs allow the production of a wide variety of display modes : ball and stick, space filling, wire-frame, ribbon drawing of protein and DNA molecules (ref. 6), dot surfaces. Stereo pairs of several models are available with MolDraw.

Parameters, such as covalent radii, atomic radii, atomic colors, atomic patterns and gray tones are adjustable.

Classical functionalities such as X, Y and Z rotations, translations, scaling, rotations around a non-ring bond are provided. An axis defined by two atoms can be moved to X, Y or Z axis. A plane defined by three atoms can be moved to XY, YZ or ZX plane. A molecule read from a crystallographic file can be projected onto a 'hkl' plane. Several molecules can be simply compared after alignment along chosen axis and planes.

Information about interatomic distances, angles or torsions are available by selecting pertinent atoms.

No energy minimization of structure is presently available. Commercial software implemented on the Macintosh II, such as CHEM3D (ref. 4) or PCMODEL (ref. 5) must be used in order to obtain this functionality.

DEPTH PERCEPTION

Several techniques have been tested for enhancing depth perception : depth-cueing, stereopsis and animation.

Depth-cueing

Depth-cueing is easily implemented on a 256 colors monitor with wire-frame, ribbon or dot surfaces models. On a black and white monitor depth-cueing requires the use of a simplification analogous to the Shademol algorithm (ref. 7). However, with space-filling or ball and stick models, hidden-surface elimination makes depth-cueing unnecessary.

Stereopsis

Stereo views are projected in either a relaxed mode, a mirror mode (one view is seen after reflection on a mirror) or a crossed mode. The stereo separation and the rotation between stereo views are adjustable. Best results are obtained after printing on a laser printer at maximum resolution.

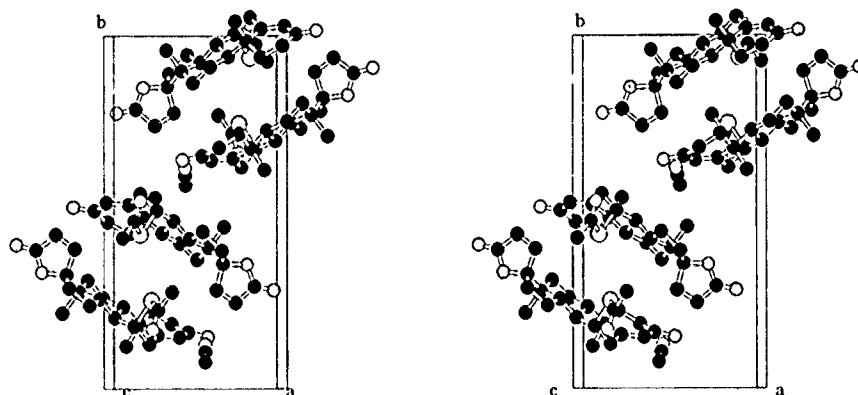


Fig. 1. Stereo pair of Spironolactone crystalized from CH_3CN (ref. 8). Details, including two CH_3CN molecules per cell are clearly visible.

Animation

Animation is a very effective technique for displaying depth relationships. Successive views of a 3-D model, projected after rotation around any axis, are stored in memory and later transferred to the screen.

As it is impossible to infer the sense of rotation from orthonormal projections of wire-frame models, animation of these models is very uncomfortable. Using depth-cued wire-frame, ball and stick or space filling models, the sense of rotation can easily be deduced.

The number of animated images and the animation speed depend on the size of the picture and on the size of available memory. About 15 color pictures covering 50%

of the screen (Macintosh II) can be transferred every second between memory and screen. Each picture uses 150 Kb of memory so that 30 pictures are stored prior to animation. For a black and white picture of the same size, transfer rate is 8 times faster and memory usage 8 times smaller.

PICTURE AND ANIMATION SAVING AND RECOVERING

Pictures and animations can be saved on disk, preferably a hard disk, to be replayed later. A typical color screen occupies 150 Kb of disk storage, a typical color animation uses, after packing, about 1.5 Mb of disk storage. Several screens and several animations can be recovered sequentially. Commercial movie makers can also be used to animate screens pasted onto the clipboard.

CONCLUSION

MolDraw and MolView programs were written to allow cheap access to many functionalities classically provided by more expensive workstations. With these programs, workstations are relieved of time-consuming picture processing and hard copies are usually more easily available than with workstations.

Contact the author for information about the availability of these programs.

REFERENCES

- 1 Inside Macintosh, Vol. I-V, Addison-Wesley Publishing Company, Inc., Reading, 1985.
- 2 Distributed by Tripos Associates Inc., 1699 South Hanley Road Suite 303, St. Louis MO 63144.
- 3 F.C. Berstein, T.F. Koetzle, G.J.B. Williams, E.F. Meyer, M.D. Brice, J.R. Rodgers, O. Kennard, T. Shimanouchi and M. J. Tasumi, The Protein Data Bank : A Computer-based Archival File for Macromolecular Structures, J. Mol. Biol., 112 (1977) 535-542.
- 4 Distributed by Cambridge Scientific Computing Inc., P.O. Box 2123, Cambridge, MA 02238.
- 5 Distributed by Serena Software, P.O. Box 3076, Bloomington, IN 47402-3076.
- 6 M. Carson, Ribbon models of macromolecules, J. Mol. Graph., 5 (1987) 103-106
- 7 M. Hahn and W.T. Wipke, SHADEMOL : An Algorithm for Presentation of Three-Dimensional Structures on a Laser Printer using Depth-Shading, Tetrahedron Comput. Method., 1 (1988) 81-86.
- 8 V. Agafonov, B. Legendre and N. Rodier, (Acta Cryst.), in press.

PRESENTATION OF GLOS

G.A. LANGLET

Commissariat à l'Energie Atomique IRDI-DESICP-DLPC-Service Chimie
Moléculaire, BP 121, 91191 GIF-SUR-YVETTE CEDEX (France)
Telex : 604641 ENERG. SPP+ ; Fax : (33) 1 69 08 79 63.

GLOS is an all-purpose software integrator that allows, in a simple way, to call numerous utilities, inter alia editors, and programs that may be written in various languages - among them : APL, Fortran, C, Pascal and assembler. It runs on PC/AT-or 80386 compatible micro-computers, and uses less memory than conventional integrators. All functionalities are accessed within pop-up menus, either with the keypad or with a mouse (e.g. IBM, Logitech, Microsoft). On-line documentation is provided ; this latter can be modified by the user when he integrates his own functionalities. GLOS requires APL*PLUS*PC (by STSC Inc., version 6 or a later version).

GLOS is ergonomic and easy to use, especially in the context of a research laboratory or a teaching environment. It has been presented at the Nancy meeting in association with BEMOL, the molecule builder and plotter (see the following description).

PRESENTATION OF BEMOL

G.A. LANGLET

Commissariat à l'Energie Atomique, IRDI-DESICP-DLPC-Service Chimie Moléculaire, BP
121, 91191 GIF-SUR-YVETTE CEDEX (France).
Télex : 604641 F ENERG. SPP+ ; Fax : (33) 1 69 08 79 63.

BEMOL is a molecular builder and plotter which has been designed for PC/AT- or 80386-compatible micros with EGA or VGA graphic card. It makes a wide use of automatic pop-up menus and of the mouse. BEMOL requires APL*PLUS*PC version 6 or up.

Only ASCII native files are used as input and/or output so as to transfer easily the data to larger machines for an eventual more complex processing.

Molecules can be built, with automatic recognition of the data type, indifferently from :

- files with Cartesian co-ordinates,
- files with oblique co-ordinates in any crystal cell,
- files containing bond lengths, bond and torsion angles,.

The format of data remains free so that, in the future, tables extracted from books or papers - e.g. Acta Cryst. reprints - can be read by a scanner. Hydrogen atoms can be automatically generated. No connectivity matrix is necessary : independent molecules and cycles are recognized. Up to 200 to 400 atoms per file - depending on the simultaneous inclusion of various other simultaneous facilities, e.g. the natural-language interpreter or the steric-hindrance checker - are handled.

Commands are accepted either from the menus or from a line that lies anywhere on the screen - and even outside, since the user's session is bufferized. In this last case, commands can be written in natural language - English or French simultaneously. The command structure has been in fact designed in order to accept the main Indo-European languages of the EEC. Command files, with natural-language instructions, can also be executed. On-line documentation is provided for the natural-language syntax.

Pictures are in 16 colours, with several stereoscopic options, including anaglyphs. Automatic framing and orientation, hidden parts, conic bonds, labelling, hardcopy are supported.

Students and scientists learn how to use BEMOL quickly and become then good teachers for their colleagues.

BEMOL has been presented at the Nancy meeting in association with GLOS, the all-purpose software integrator (see the preceeding description).

NETWORK COMPUTING OF THE EPSTEIN-NESBET SECOND-ORDER CORRECTION TO THE MOLECULAR ELECTRONIC DENSITY. PRELIMINARY RESULTS.

M. PICARD and J.M. LECLERCQ

Laboratoire de Dynamique des Interactions Moléculaires*, Université Pierre et Marie
Curie, Campus Jussieu, tour 22, 75252 Paris Cedex 05 (France)

SUMMARY

Parallel processing, in the environment of an ETHERNET local area network of UNIX-based workstations, of the Epstein-Nesbet second-order correction to the density matrix is outlined. The speed-up is reported for preliminary investigations of the electronic ground state of three testing compounds: H_2O , H_2CO and NH_2NO_2 . These first results support the point of view that Network Computing is an efficient and low-cost tool for theoretical investigations of molecular properties.

INTRODUCTION

It is well known that parallel processing has been introduced because improvements in electronic circuit speeds alone cannot produce the performance required by many problems (refs. 1,2 ; of the two techniques for introducing parallelism, i.e. replication and pipelining, we emphasize the first one in the present note). The CRAY 2 and IBM 3090/600 E(S)-VF computers are typical examples of multi-processor mainframes which offer parallelism capability. Such an approach of the parallelism, based on the limited replication of very fast processors is very expensive but well-adapted to a large scale of problems from the weather forecast to military applications... The use of mass-produced VLSI circuits in massive replication architectures is an alternative technique which often requires drastic adaptations of the usual algorithms. The Network Computing, i.e. the parallel processing in the environment of a network of workstations and servers, is an intermediate technique with limited replications of mass-produced VLSI circuits. In such a case, the optimization of the "cost / performance" ratio (not the power at any cost !) is the priority.

Among the methodologies of Quantum Chemistry, the perturbation expansion (refs. 3,4) is particularly adapted to parallel processing. However, as far as we know, the first investigations (refs. 5,6 and references therein) are limited to the calculations of the electronic correlation energy. We focus ourselves on the calculation, beyond the SCF level, of the molecular electronic density (refs. 7-9) and the present note reports developments of the Epstein-Nesbet (refs. 10-12) correction to the matrix density elements.

*UPR A0271, CNRS, France;

Unité liée par Convention à l'Université Pierre et Marie Curie, Paris, France.

THE EPSTEIN-NESBET SECOND-ORDER CORRECTION TO THE DENSITY MATRIX.

The Möller-Plesset partition is defined as (refs. 13,12) :

$$H_{\text{exact}} = H^{(0)} + V = \sum_i F(i) + V \quad (1)$$

where F is the closed-shell restricted Hartree-Fock operator and \sum_i stands for the summation over the N electrons of the molecular system, while the Epstein-Nesbet partition is:

$$H_{\text{exact}} = H^{(0)} + V = \sum_i F(i) + \sum_K \langle K | V | K \rangle | K \rangle \langle K | + V \quad (2)$$

where \sum_K stands for the summation over all the eigenfunctions of $H^{(0)}$ (and $H^{(0)}$):

$$|K\rangle = \frac{1}{\sqrt{N!}} \sum_P (-1)^P P \{ \phi_{K1}(r_1) \sigma_{K1}(s_1) \phi_{K2}(r_2) \sigma_{K2}(s_2) \dots \phi_{KN}(r_N) \sigma_{KN}(s_N) \} \quad (3)$$

where the molecular orbitals (MO) ϕ_{Ki} , issued from the SCF calculation, are expanded on atomic orbitals (AO) χ_μ as:

$$\phi_{Ki} = \sum_\mu d_{\mu,Ki} \chi_\mu \quad (4)$$

While the theoretical investigation of the electronic density at the SCF level is very popular, as far as we know, the corresponding investigation beyond this step is till now limited (refs. 14,15 and references therein). In the framework of the perturbation expansion, the second-order correction to the average value, for the electronic ground state, of the density operator ρ is (on the assumption of real basis sets):

$$\rho_{00}^{(2)} = 2 \int \Psi_0^{(0)} \rho \Psi_0^{(2)} d\tau + \int \Psi_0^{(1)} \rho \Psi_0^{(1)} d\tau \quad (5)$$

where

- $\rho(r) = \sum_i \delta(r_i - r)$ ($\delta(r_i - r)$ is the usual Dirac function);
- $\Psi_0^{(0)}$, $\Psi_0^{(1)}$ and $\Psi_0^{(2)}$ are the zeroth-order wave function of the reference state and its first- and second-order corrections, respectively.

On the basis of the classical expressions of $\Psi_0^{(1)}$ and $\Psi_0^{(2)}$, the Epstein-Nesbet second-order correction to the matrix density elements may be easily written as:

$$\begin{aligned}
P_{\mu\mu}^{(2)} = & - \left[\sum_{l \in \{2/0\}} \left[\frac{V_{l0}}{E_0^{(0)} - E_l^{(0)}} \right]^2 \right] \left\{ 2 \sum_{i \in 0} d_{\mu,0i}^2 \right\} \\
& + 2 \sum_{K \in \{1/0\}} \left\{ \left[\sum_{l \in \{2/0\} \text{ and } \epsilon \in \{2/K\}} \frac{V_{Kl} V_{l0}}{(E_0^{(0)} - E_l^{(0)}) (E_0^{(0)} - E_K^{(0)})} \right] \{ d_{\mu,0K} d_{\mu,K\mu} \} \right\} \\
& + \sum_{K \in \{2/0\}} \left[\frac{V_{K0}}{E_0^{(0)} - E_K^{(0)}} \right]^2 \left\{ \sum_{i \in K} n_{Ki} d_{\mu,Ki}^2 \right\} \\
& + 2 \sum_{K \in \{2/0\}} \sum_{L \in \{2/0\} \text{ and } \epsilon \in \{1/K\}}^{<K} \left[\frac{V_{K0}}{E_0^{(0)} - E_K^{(0)}} \right] \left[\frac{V_{L0}}{E_0^{(0)} - E_L^{(0)}} \right] \{ \epsilon_{KL} d_{\mu,Ki} d_{\mu,Lj} \}
\end{aligned} \tag{6}$$

$$\begin{aligned}
P_{\mu\nu}^{(2)} = & - \left[\sum_{l \in \{2/0\}} \left[\frac{V_{l0}}{E_0^{(0)} - E_l^{(0)}} \right]^2 \right] \left\{ 2 \sum_{i \in 0} d_{\mu,0i} d_{\nu,0i} \right\} \\
& + 2 \sum_{K \in \{1/0\}} \left\{ \left[\sum_{l \in \{2/0\} \text{ and } \epsilon \in \{2/K\}} \frac{V_{Kl} V_{l0}}{(E_0^{(0)} - E_l^{(0)}) (E_0^{(0)} - E_K^{(0)})} \right] \{ d_{\mu,0K} d_{\nu,K\nu} \} \right\} \\
& + \sum_{K \in \{2/0\}} \left[\frac{V_{K0}}{E_0^{(0)} - E_K^{(0)}} \right]^2 \left\{ \sum_{i \in K} n_{Ki} d_{\mu,Ki} d_{\nu,Ki} \right\} \\
& + 2 \sum_{K \in \{2/0\}} \sum_{L \in \{2/0\} \text{ and } \epsilon \in \{1/K\}}^{<K} \left[\frac{V_{K0}}{E_0^{(0)} - E_K^{(0)}} \right] \left[\frac{V_{L0}}{E_0^{(0)} - E_L^{(0)}} \right] \{ \epsilon_{KL} d_{\mu,Ki} d_{\nu,Lj} \}
\end{aligned} \tag{6'}$$

where:

- 1/0, 2/0 stand for single and double excitations with respect to the zeroth-order wave function of the electronic ground state (an unique closed-shell Slater determinant, according to our hypothesis; hereafter 0), respectively;
- l, K, L are eigenfunctions of $H^{(0)}$, $E_l^{(0)}$, $E_K^{(0)}$, $E_L^{(0)}$ the corresponding eigenvalues, V_{l0} , V_{K0} , V_{L0} the corresponding matrix elements of V with the "ket" 0;
- 1/K, 2/K stand for single and double excitations with respect to a given Slater determinant K (which arises itself from a single or a double excitation with respect to the zeroth-order wave function of the ground state), respectively;
- n_{Ki} is the occupation number of the MO i in K;
- $\epsilon_{KL} \pm 1$ following the parity of the permutation which places at the same location the identical spin-orbitals of K and L.

CALCULATIONS AND RESULTS

The electronic ab initio calculations have been carried out using the MONSTER-GAUSS program (ref. 16). The SCF process within the closed-shell restricted Hartree-Fock formalism has been performed using a 6-31G**, a 6-31G* or a 6-31G basis set for the electronic ground state of H_2O , H_2CO , H_2NNO_2 , respectively (in order to have 25 to 40 basis set functions at this step of preliminary testing calculations). The integral transformation from AO to MO basis has involved all the MO (no frozen occupied or virtual MO). SCF calculations and integral transformations run on the SUN 3 / 260 workstation (see Fig. 1).

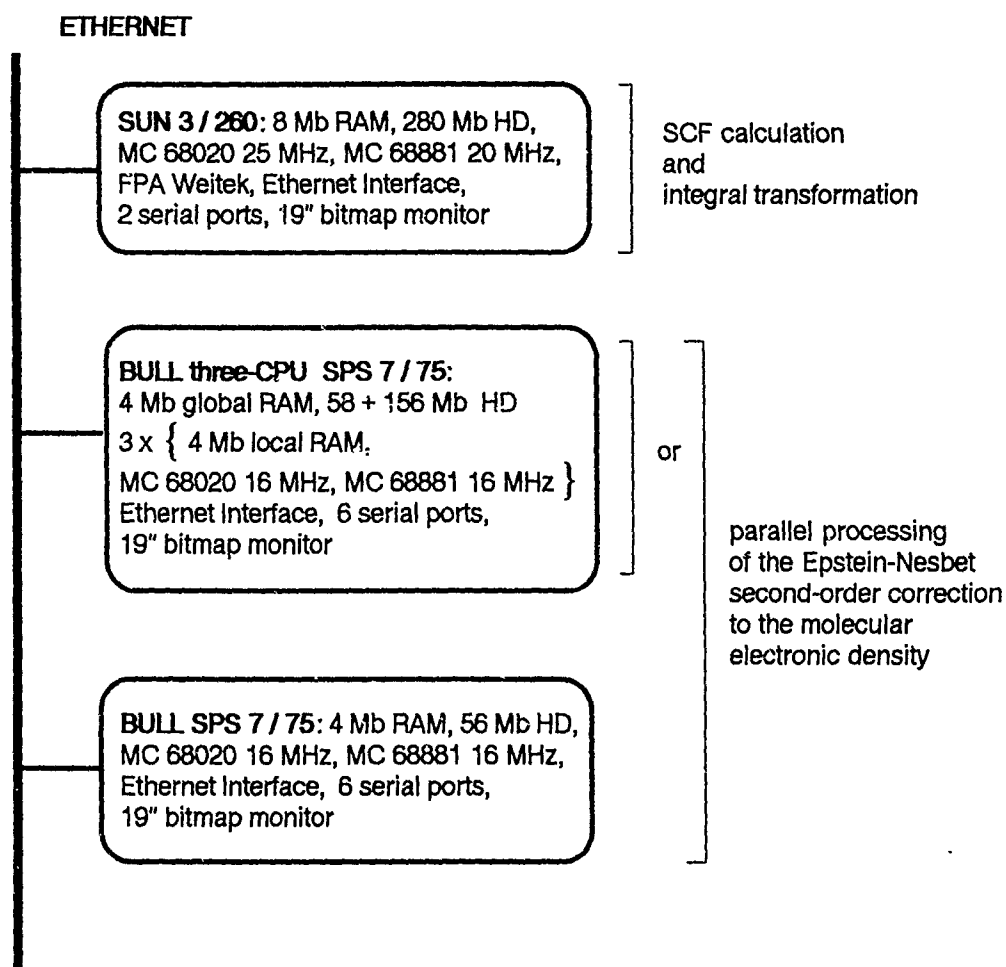


Fig. 1. The three workstations of the local area network used in our calculations.

The unformatted direct-access output file fort.8 of the SCF process has been saved and transferred to the SPS 7 workstation (s) with the usual remote copy order of ETHERNET protocol, in order to dispose of the MO vectors (coefficients) $d_{\mu,Ki}$ and the overlap integrals on the AO basis (for the population analysis). Similarly, the unformatted sequential-access output file fort.10, resulting of the integral transformation, has been saved and transferred to the SPS 7 workstation (s) in order to dispose of the transformed one- and two-electron integrals (at this step, it may be pointed out the possibility to use unformatted files from a workstation to another one, due to their MC 68XXX-based character).

The Epstein-Nesbet second-order correction to the matrix density has been calculated on the three CPUs of the multiprocessor BULL SPS 7 workstation alone or on this workstation and the one-CPU SPS 7 together (technical considerations are available on request; see also ref. 6).

The usual parameter that is used to evaluate the performance of parallel calculations is the speed-up $S_p = T_1/T_p$, where T_1 and T_p are the times for the algorithm to run on one CPU (sequential calculation) or p CPUs (parallel calculation), respectively. The values of this parameter are reported in Table 1 for three testing investigations.

TABLE 1

The speed-up S_p for some preliminary testing calculations.

Testing calculation	speed-up S_p	
	3-CPU SPS 7 alone ($p=3$)	3-CPU plus 1-CPU SPS 7 ($p=4$)
H ₂ O / 6-31G**	2.89	3.81
H ₂ CO / 6-31G*	2.90	3.83
H ₂ NNO ₂ / 6-31G	2.92	3.87

These values of the speed-up show how the present Network Computing is well-adapted to the second-order correction to matrix density elements (the theoretical limits of S_p are 3 and 4 for $p=3$ and 4, respectively). On another hand, the excellent "cost / performance" ratio of mass-produced microprocessors, which leads to the well-known "downsizing" in a large scale of applications, ensures the low cost of such investigations. Consequently, we can claim that these first results support the point of view that Network Computing is an efficient and inexpensive tool for theoretical studies of molecular properties.

REFERENCES

- 1 D.J. Kuck, The Structure of Computers and Computations, Vol. 1, John Wiley and Sons, New York, 1978.
- 2 R.W. Hockney and C.R. Jesshope, Parallel Computers 2, Adam Hilger, Bristol and Philadelphia, 1988.
- 3 J.-L. Rivail, Éléments de Chimie Quantique à l'usage des chimistes, InterEditions-/Editions du C N R S, Paris, 1989.
- 4 L. Landau and E. Lifchitz, Mécanique Quantique Théorie non Relativiste, Editions Mir, Moscou, 1966.
- 5 J.D. Watts and M. Dupuis, Parallel Computation of the Möller-Plesset Second-Order Contribution to the Electronic Correlation Energy, J. Comp. Chemistry, 9 (2) (1988) 158-170.
- 6 M. Bendrider, rapport de stage, DEA de Chimie Informatique et Théorique, Université de Paris-Sud (France), juillet 1989.
- 7 R. McWeeny and B.T. Sutcliffe, Methods of Molecular Quantum Mechanics, Academic Press, London and New York, 1969.
- 8 R.S. Mulliken, Electronic Population Analysis on LCAO-MO Molecular Wave Functions. I, J. Chem. Phys., 23 (10) (1955) 1833-1840; II. Overlap Populations, Bond Orders, and Covalent Bond Energies, *ibid* 1841-1846; III. Effects of Hybridization on Overlap and Gross AO Populations, J. Chem. Phys. 23 (12) 2338-2342; IV. Bonding and Antibonding in LCAO and Valence-Bond Theories, *ibid* 2343-2346.
- 9 S. Fliszar, Charge Distributions and Chemical Effects, Springer-Verlag, New York, 1983.
- 10 P.S. Epstein, The Stark Effect from the Point of View of Schrödinger's Quantum Theory, Phys. Rev., 28 (1926) 695-710.
- 11 R.K. Nesbet, Configuration Interaction in Orbital Theories, Proc. Royal Society, A 230 (1955) 312-321.
- 12 S. Diner, J.-P. Malrieu, and P. Claverie, The Use of Perturbation Methods for the Study of the Effects of Configuration Interaction II. Variation of the 2nd Order Energy Correction in the Series of Linear Polyenes and Polyacenes, Theoret. Chim. Acta, 8 (1967) 390-403.
- 13 C. Möller and M.S. Plesset, Note on an Approximation Treatment for Many-Electron Systems, Phys. Rev., 46 (1934) 618-622.
- 14 S. Fliszar, J.M. Leclercq, C. Mijoule, and S. Odier, Alkane Atomic Charges in Energy Calculations, in: V.H. Smith, Jr., H.F. Schaefer III, and K. Morokuma (Eds.), Applied Quantum Chemistry, D. Reidel Publishing Company, Dordrecht, 1986, pp. 395-401.
- 15 C. Mijoule, J.M. Leclercq, M. Comeau, and S. Fliszar, Density Matrix Contribution to the Charge Distribution in Hydrocarbons, Arising from Singly Excited Configurations in CI Calculations, submitted for publication.
- 16 MONSTERGAUSS Program, M.R. Peterson and R.A. Poirier, Department of Chemistry, University of Toronto, Toronto, Canada.

COMPUTATIONAL CHEMISTRY ON SUPERCOMPUTERS

François SCHWAAB
CRIN and INRIA
BP 239
54506 VANDŒUVRE LES NANCY Cedex

This paper summarizes the introduction to the panel discussion about Supercomputing in Chemistry. It presents the State of the Art in Supercomputing, the new wave constituted by massively parallel architectures, the new trends and a glossary of technical terms.

Introduction :

To introduce the problem we report some quotations.

Klaus Schulten, University of Illinois, Urbana Champaign.

" The supercomputer is like an intelligent microscope. It allows you to look into the protein, to magnify it, and to slow down its motion"

Jacob V. MAIZEL, National Cancer Institute, Laboratory of Mathematical Biology.

" Using the supercomputer for computational simulations of biochemical systems can both eliminate experimental work on unpromising drugs candidates and provide insights leading to new approaches."

David A. DIXON, Du Pont.

" The combination of high-speed computers, theoretical methods, and software now makes it possible to perform simulations of complex systems and processes of real commercial interest to the chemical industry."

The State of the Art in Supercomputer Architecture :

1. A short review of supercomputer's history :

- 1960 : Univac and CDC build the first supercomputers. IBM followed with its 360/91 model. The well known CDC 6600 is the best example of this generation of supercomputers, it was designed by Seymour CRAY.
- 1969 : CDC builds the 7600.
- 1976 : CRAY 1 : the first vector processing supercomputer.
- 1981 : CDC wants to compete CRAY with the CYBER 205.
- 1982 : CRAY XMP/2 : Cray boosts the vector processing with parallelism.

- 1985 : CRAY 2 with large memory size. The Japanese enter the competition : FUJITSU VP 200, HITACHI S810 and NEC SX1-SX2.
- 1986 : IBM builds the 3090 with Vector Facility.
- 1987 : CDC builds the ETA-10 serie.
- 1988 : CRAY launches the YMP, HITACHI the S820 and ETA disappears !
- 1989 : New supercomputers are announced :
 - FUJITSU VP2000,
 - NEC SX-X,
 - SUPERCOMPUTER SS 1,
 - CRAY 3 : new technology (GaAs), 16 processors each 100 times faster than the CRAY 1.

2. Some data representing the situation of Supercomputers at the present time.

The size of the market and distribution of supercomputers makes :

400 Supercomputers installed in the world : Cray 56%, Fujitsu 20%, Nec 6%, etc.
In Europe : 58 Cray, 12 CDC, 7 Siemens-Fujitsu, etc.

Present day performances :

- Cray YMP-8 : 1.6 GFlops,
- Nec SX-3 : 680 Mips and 22 GFlops (announced),
- Fujitsu VP2000 : 4 GFlops.

Which Operating System ?

UNIX® becomes the standard supercomputer Operating System and Local Area Networks combine supercomputers and graphic workstations. Currently there is a new trend : hierarchical computing.

3. Important comment :

With more and more power available, it is becoming increasingly critical to pay attention to efficient programming on a Supercomputer which needs good knowledge of :

- Supercomputer architecture,
- Vectorization techniques,
- Program parallelization with loop splitting.

This avenue of research constitutes a major objective for the companies manufacturing Supercomputers.

The New Wave : Massively Parallel Computers

With the limitations of technology, the principal direction of development to increase computation speed is the exploitation of parallel architectures.

® UNIX is a trade mark of AT&T.

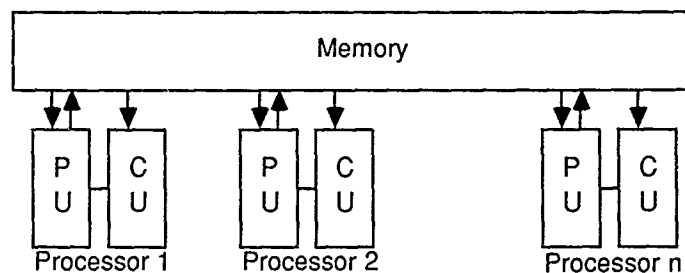
In the early '80s a new concept appeared, namely the network of tightly coupled processors.

1. First, recall the Michael FLYNN classification :

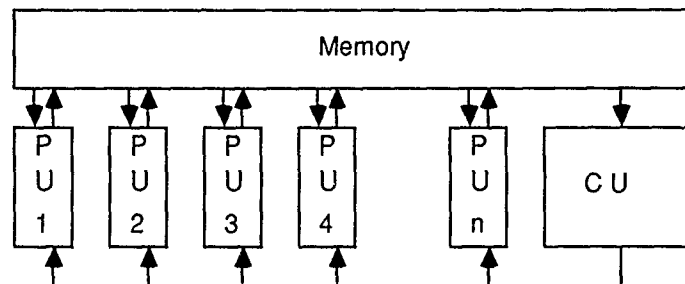
- MIMD : Multiple Instruction Multiple Data.
- SIMD : Single Instruction Multiple Data.

The main difference between these two architectures is the number of Control Units. In a MIMD computer, each Processing Unit is associated with a Control Unit whereas in a SIMD computer there is only one single Control Unit for all the Processing Units.

MIMD



SIMD



P U : Processing Unit
C U : Control Unit

Diagram 1 : Comparison between the MIMD and the SIMD Architecture model.

2. Granularity : In such architectures the choice of 'size' of the processing element has to be made.

This 'size' is referred to as granularity :

In a fine grain architecture, each processing element manipulates a small set of bits with elementary operations whereas in a coarse grain architecture, each processing element is an entirely complete processor mostly working on 32 bit data.

Usually fine granularity is associated with a SIMD approach, whereas coarse granularity and MIMD are linked.

3. The communication network plays an important rôle because massively parallel processing needs high data exchange rate between the processors. The network interconnects the processors according to one of the following models :

- (i) - Mesh Array,
- (ii) - Perfect Shuffle,
- (iii) - Hypercube,
- (iv) - Pyramid.

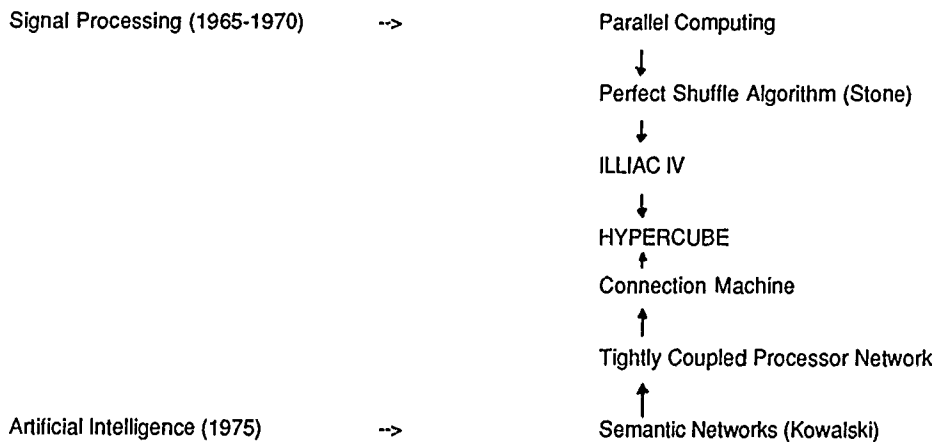
It is important to balance the cost of data exchange between the processors versus the gain in parallel processing.

N.B. (ii) and (iii) represent identical environnements but whereas the Perfect Shuffle is dynamic the Hypercube is static.

4. The application fields :

The numerous fields of application of Supercomputing in number crunching, Supercomputing in mathematics and Signal processing are of obvious importance. It is also useful for pattern matching in data bases exploiting the associative memory feature and for image processing and Ray Tracing, and last, but not least in Artificial Intelligence, Expert systems and Pattern Recognition.

5. The historical steps which led to the design of the Connection Machine :



A short Review of Massively Parallel Computers :

A number of companies each offer a product of this type :

The Butterfly Computer for Signal Processing made by BBN Advanced Computer Inc. It is a MIMD architecture with 256 coarse grain processors. Each processor is built up of a Motorola 68020 with Floating Point co-processor and 4Mbyte local memory. Data exchange is achieved by routing processors and the network links the processors according to the Perfect Shuffle model.

The Connection Machine made by Thinking Machines Corporation :

This is a SIMD architecture built up of a dimension 16 Hypercube comprising 65536 fine grain processors. Each processor manipulates only 1 bit and it has 4Kbit local memory. The peak performance is 12 GIPS and the mean is 2 GIPS.

The hardware organization follows this hierarchical model : one chip contains 16 processors interconnected by bus, one board contains 32 chips and the machine contains 128 boards. The 4096 chips of the Hypercube computer are interconnected by a routing processor.

Intel Scientific Computers produce the IPSC.

This is MIMD architecture constituted by a dimension 7 Hypercube with 128 coarse grain processors. Each processor being either a 80286 and 80287 for the IPSC or 80386 and 80387 for the IPSC/2. Various configurations with 1, 4, 8 or 16 Mbytes local memory with cache are available. The network is controlled by a direct message router. The performances for the IPSC are 100 MIPS and 8 MFLOPS whereas for the IPSC/2 they are 1 GIPS, 150 MFLOPS and 1 GFLOP with vector facility.

The Transputer by INMOS Limited :

The Transputer is a brick to build up MIMD massively parallel computers. The Transputer T800 processor is coarse grain with a 32 bit RISC architecture with 4 Kbytes local memory and 4 Gbytes external memory. The network's interconnection uses 4 links at 20 Mbit/s to communicate. The performances are 15 MIPS RISC and 2 MFLOPS. An original feature is the programming language called OCCAM which is a parallel language based on the CSP model (Hoare).

Two examples of existing Transputer based Supercomputers are the Computing Surface from Meiko and the T-Node from Telmat. Both architectures are highly modular with 4 to 1024 processors and up to 4 Gbytes memory. In addition, the T-Node offers a dynamically reconfigurable network.

A glossary of technical terms appropriate to the field :

- PIPELINING :

Cutting up the interpretation algorithm of the instruction set in steps which are sequentially executed.

The main problems are the flushing of the pipeline by conditional branches and the memory becoming the bottleneck.

- MEMORY MANAGEMENT :

- Interleaved Banks :

To parallelize access to contiguous elements, the main memory is split into 2^n independant banks. Two contiguous addresses map in two different banks and addresses in one bank progress with a 2^n step.

- Cache Memory :

Small but high speed memory managed according to a LRU policy. Only the Least Recently Used information is stored.

- VECTOR PROCESSING :

A set of high speed vector registers is available to the processor. The processing of these vector registers is carried out by special operators in a pipelined mode, the main problem being the dependencies between the elements of the vectors which cause the flushing of the pipeline.

- AMDAHL'S LAW :

This law predicts the performance of a machine that has 2 separate modes of computing (scalar and vector) as a function of the fraction of computation devoted to each mode.

- RISC : Reduced Instruction Set Computer

The main idea is to reduce and optimize the internal architecture of the processor by optimizing the most frequently used instruction using pipelining and enhancement of the set of internal high speed registers and simplifying access to the memory.

NEW TRENDS TOWARDS ARTIFICIAL NEURAL NETWORKS :

The main goal is to mimic human thinking for :

- (i) - Pattern classification,
- (ii) - Speech recognition,
- (iii) - Vision processing.

The method adopted is to simulate nerve cells, their interconnections and pattern of interaction.

The processing element receives signals from neighbouring elements and, according to a system of weighting factors, determines whether to pass the signal farther up the network or not.

The basic neural network properties are massive parallelism and adaptivity, (neurons work collectively), fault tolerance, training rather than programming, (the weighting factors are computed through learning) and data processing by spreading activation from input to output.

The neural network works as a Dynamic System controlled by a transfer function with natural functioning like an associative memory.

The learning algorithm uses error back propagation : the difference between the expected and the obtained result is computed and the weights are corrected.

The domains of application are optimization problems with function minimization and pattern classification.

State of the Art : currently most work consists in simulation on supercomputers, because the learning algorithm requires high speed computers, and specialised hardware design based on analog or digital integrated circuits. Today we can fit 100 neurons on 1 chip and in 50 years perhaps computers with 10^{10} neurons.

HIGH SPEED GaAs TECHNOLOGY :

The main properties of this technology are high electron mobility with high peak electron velocity and improved radiation tolerance over a wide range of operating temperature. GaAs reduces power consumption and improves speed : in the fact GaAs is 5 times faster than Si ECL technology.

State of the art :

Today GaAs VLSI integrated circuits contain for 10^3 to 10^5 gates whereas silicon VLSI integrated circuits contain 10^7 gates. GaAs will never totally replace Silicon, it will be used in selected aerospace, defense and supercomputer applications.

Conclusion :

Supercomputers are now 30 years old ... I think we are in the middle-ages !

Acknowledgements :

I am grateful to D. TUSERA of the INRIA for his help in compiling the bibliography. The text in English was reviewed by Ph. HOGGAN.

Bibliography :

V. Milutinovic : GaAs Microprocessor Technology. Computer October 1986

L. E. Larson et al. : GaAs High-Speed Digital IC Technology : an overview. Computer October 1986

H. S. Stone : Parallel Processing with the Perfect Shuffle. Transactions on computers, vol C-20, n°2, February 1971

W. D. Hillis : The Connection Machine. MIT Press 1985

S. E. Fahlman et al. : Connectionist Architectures for Artificial Intelligence. Computer January 1987

O. Lubeck et al. : A Benchmark Comparison of Three Supercomputers : Fujitsu VP-200, Hitachi S810/20 and Cray X-MP/2. Computer December 1985

DISCUSSION

The discussion started with a short exposé by Messrs. Charles Henriët (Cray Research France), Armand Herscovici (IBM France), Jean-Claude Reynaud (Convex S.A.) and Daniel Urbain (FPS Computing). The following salient points have been recorded.

Ch. Henriët

Exciting new developments of theoretical and computational methods, combined with major advances in computer hardware and software will continue to make computational chemistry one of the most fascinating fields of scientific computing. Progress in computational chemistry continues to increase our ability to tackle more complex problems at a higher level of accuracy than in the past. The most important opportunities for computational chemistry appear in molecular biology, including pharmaceutical, agrochemical, and biotechnology research, and materials sciences, especially polymer sciences, catalysis, and advanced electronic, optical, and structural materials. We have reached a threshold where, perhaps for the first time in chemical research, large-scale numerical simulations are becoming an important element in industrial research. Given this emerging economic relevance of computational chemistry and the continued progress in hardware and software, the future of computational chemistry looks extremely promising. Three major technological factors will influence the future of the field : new theoretical/computational methods, new hardware, and new software tools.

Future theoretical developments

The most fundamental factor is the development of new computational theories and approaches. During the past decade, single-reference Hartree-Fock methods with the ability to calculate first and second energy derivatives analytically have become a standard tool in theoretical organic chemistry and form the backbone in the development of force-field parameters for molecular dynamics studies of biomacromolecules and polymers. In traditional implementations of *ab initio* methods integrals were stored externally. In such an approach, the N^4 scaling problem constitutes a serious roadblock for the treatment of systems with more than about 200 basis functions. Direct SCF methods, now available in programs such as Gaussian-88, are circumventing this problem and significantly larger systems can be tackled. As it turns out, for systems with more than several hundred basis functions, the scaling reduces from N^4 to N^3 and even below. Hence, increased computer power and the implementation of parallel algorithms open up the possibility to treat much larger systems on the *ab-initio* level than previously thought possible.

We can also expect a continued progress in high-precision correlated methods. In fact, for small systems, it will become possible to carry out full CI calculations more and more readily thus establishing benchmarks on which various basis set truncations and limitations in the CI expansion can be systematically evaluated. As correlated methods are applied to larger and larger systems, it can be expected that localization of correlation could keep the high scaling factor of correlated methods within reasonable bounds, similar to what we are observing for the single-reference methods.

Two newer developments in the treatment of the many-body problem can be anticipated to bear fruit in computational chemistry. One is density functional theory, where programs are

now being developed that contain analytic first derivatives. This capability will allow the systematic study of structural properties of a wide variety of molecules, clusters, solids, and surfaces including transition metals, rare earths, and actinides thus expanding the scope of systems that can be treated from first principles. Today, the most common implementation of density functional theory is the so-called local density approximation (LDA), yet several ways are currently being explored to go beyond this approximation. A close collaboration between high-precision CI or multireference practitioners and developers of density functional theory should prove extremely fruitful as we strive to develop more accurate and efficient methods for the challenging many-body problem of correlated electron systems. Other interesting new theoretical development are the quantum Monte-Carlo methods. Although slow in coming, there are promising signs that by using pseudo-Hamiltonians these methods could be successfully applied to larger systems and avoiding the challenging computational problems in this approach.

The combination of electronic structure theory with molecular dynamics can be anticipated as one of the most exciting future developments in computational chemistry as this aims directly at the simulation of dynamical phenomena of the making and breaking of chemical bonds. One can envision a combination of quantum mechanics and force field methods where active sites of an enzyme or reactive centers of a catalyst are treated quantum mechanically whereas the surroundings are described by a quasi-classical force field. Another, more profound combination between quantum mechanics and molecular dynamics is possible in the Car-Parrinello scheme in which the motion of atoms and changes in the wave-functions are conceptually treated on the same footing.

Future hardware - network supercomputing in heterogeneous environments

During the next decade, progress in computer hardware will be dominated by parallelism. While the performance of single processors will be pushed by further reducing the clock-cycle perhaps to 1 ns by the late 1990s, major performance enhancements will come from parallel processing and from large-memory architectures. Although certain tasks in computational chemistry are intrinsically perfectly parallel such as the evaluation of two-electron integrals in *ab-initio* methods or the pair-wise energy evaluations in molecular dynamics, it would be a mistake to assume that massively parallel architectures will be the panacea of computing. There are steps in all large-scale simulations that are best performed in a shared-memory architecture. Perhaps the proper balance between powerful scalar and vector, shared-memory processors at a moderate level of parallelism connected to highly parallel, distributed memory processors in the network will provide the best solution for large-scale numerical tasks in chemistry.

We believe that the intimate connection between desk-top graphics workstations with supercomputers and large data-base systems has the potential for an extraordinarily productive environment. The key is connectivity between heterogeneous hardware which will be enabled by software standards. It appears that UNIX¹ is becoming such a desired standard, perhaps initially with TCP/IP as network protocol.

Since the bandwidth of networks can be expected to remain the critical bottleneck in a network-supercomputing environment, sufficient local and desk-top computer power will be essential for adequate load balancing and data compression over the network.

Future software languages and tools

During the past two decades, FORTRAN was the standard for scientific programming. There is an incredible wealth of chemistry software written in this language, with an investment in man-

¹ UNIX is a trademark of AT&T.

time that is probably measured in thousands of man-years. Concurrently with the software developments, major progress has been made in FORTRAN compiler technology and code optimization techniques which also represent an enormous economic investment. As we move to new software environments, it will be wise to design them in such a way that the earlier programming work can be used. For example, as more and more software is being written in C, this compatibility can be achieved since C and FORTRAN coexist quite well. In fact, the trend seems to be that the computationally intensive parts will be written in FORTRAN whereas the program-system interfaces such as connections to graphics workstations and the handling of data-transfer across networks will be expressed in C using the obvious relationships between C and UNIX.

Today, parallelism of FORTRAN codes is usually handled via compiler directives that are inserted automatically by pre-processors or manually by the programmer. We will have to wait and see which standards will emerge for expressing parallelism within existing computer languages or perhaps within new languages. Perhaps the key issue here is not so much if we need yet another language on the same level as FORTRAN, C, or PASCAL, but rather the creation of higher levels of language for expressing mathematical and chemical concepts.

An important trend in computational chemistry will be the embedding of numerically intensive tasks in an environment that allows the convenient manipulation of graphical, symbolic and logical information. Such an environment will enable the research chemist to combine in a productive fashion information from experience and experiment with computed data from simulations. Expert-systems may assist the non-expert to carry out meaningful calculations, guiding in the choice of the computational approach, and assisting in the interpretation and assessment of results. Nevertheless, these systems will not replace the need for more education of practicing and future chemists. In fact, education is clearly the most critical component of all to ensure a prosperous future of computational chemistry. A wider group of chemists need to be taught how to use all the powerful tools of computational chemistry.

In conclusion, we are excited about many future opportunities of computational chemistry as applied to the fascinating areas of the life-sciences as well as materials sciences. Hopefully, continued advances in computational chemistry will make this branch of science a widely used discipline and true partner to the experimental efforts as we strive to solve the many challenging problems such as cures for cancer and AIDS, the development of novel electronic, optical, magnetic, and structural materials, and the development of safer chemical processes and compounds that are environmentally sound.

A. Herscovici

Numerical intensive computing (NIC) spread widely during the 1970's, when the petroleum problem generated a worldwide economic crisis, with an emerging tough competitive environment. To meet these new conditions, companies had to implement additive components in their strategies.

One of them consisted in organizing their R&D departments in order to enable them to issue quicker more and more optimized new products, to face competition efficiently. This meant shorter development cycles, and more discrete simulations, both items driving to more and more powerful computers. We can consider this to be the very origin of the strong growth of the NIC market, estimated above 30% per year worldwide.

Research organization fully participated in this scheme, as most advanced applications, which feed supercomputers, are conceived within the research world.

Calculations generated by this new environment were huge, compared with the computing power available. This resulted in enormous CPU time, ending in the main concern : increase

computing power in order to shorten the NIC process. Most people were indirectly considering that increasing computer speed, and shortening the development cycle, were two similar concepts.

Emphasis was given to supercomputers, and vector speed : the bigger the vector peak rate, the more efficient the supercomputer from a development cycle point of view.

Very quickly, supercomputer users (scientists, engineers) realised that applications could be vectorised only up to a certain amount, and that the CPU performance was a combination of scalar and vector speed, scalar speed being a major contributor. The concept of the balanced supercomputer was born.

Supercomputer vendors provided more and more powerful machines, as technological advances were coming fast. Users immediately took advantage of the available computing power, generating new CPU bound calculations, and models extraordinarily increased in size.

The result was dramatic and rather unexpected : the nature of NIC applications was definitively changed.

To analyse this very deep change, let's go back 20 years ago, when NIC applications could be analysed in a few words by the following spectrum : far above 90% in CPU, far below 10% in I/O. As a consequence, computers for NIC had to emphasize CPU speed only.

Nowadays, this scheme is definitively obsolete . NIC applications are using large quantities of data, and producing a huge amount of results : pre and post processing are taking on tremendous importance , storage size problems appear , I/O management becomes a major factor for performance ; true cooperative processing is mandatory.

Moreover, organisation problems appear, which were unknown up to now . designing a finite element mesh may take several months (for instance in automotive industry in car body simulation). Companies can no longer afford to throw away this mesh after the calculation is done . there must be some way to save it in a technical data base, for later re-use in another version of the body. This applies also to the enormous amount of results.

In addition, similar runs by different persons must be avoided. This is not an easy matter to deal with, as many NIC organizations have not yet been able to take into account the dramatic change in NIC applications, due to the quickness of the phenomenon . most are still working within an organization filling the former spectrum, ending in a partial loss of control . very often, more than 50% of the runs could have been avoided. A DP organization taking advantage of the technical data base has to be implemented, ensuring full control.

Another major issue is the fact that supercomputers are large expensive systems, so that management wishes the machine be used simultaneously by many people. This drives to a new requirement, rather seldom taken into account in NIC . optimize the use of the system, which is an operating system problem. The performance of the computer is no longer to be measured on a kernel, or a single application, but on a set of representative programs run simultaneously . the machine with the best performance will produce the results for ALL the programmes in the shortest amount of time.

To-day, shortening the research or development cycle implies taking into account all these new parameters . NIC is becoming mature when it really meets these requirements. This is far wider than concentrating on CPU performance.

J.C. Reynaud

This speaker emphasized the prosperous future of vectorized superminicomputers which have a very attractive performance/cost ratio due to the fact that they are air-cooled. He also insisted on the important contribution to homogeneity introduced by the use of the UNIX operating system, that brings a new freedom to the users for choosing the best hardware platform without being tied up by a proprietary operating system. He concluded in saying : «UNIX is a challenge for the computer manufacturers. Besides, to maintain our leadership in air-cooled supercomputers we develop a products strategy based on advanced technology as for instance CMOS VLSI, ECL and GaAs on future generation in order to multiply by 10 the machine performances every 3 years.»

D. Urbain

M. Urbain agreed with the importance of the UNIX operating system and of the vectorized supermini, with a special attention paid to the RISC architecture and the role of a good graphics display.